

3c715 U.S. PTO  
07/19/00

Please type a plus sign (+) inside this box → **+**

PTO/SB/05 (4/98)  
Approved for use through 09/30/2000. OMB 0651-0032  
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE  
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

# UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 C.F.R. § 1.53(b))

Attorney Docket No. **114465.401**

First Inventor or Application Identifier **Thastrup et al.**

Title **Novel Fluorescent Proteins**

Express Mail Label No.

## APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. ☒ \* Fee Transmittal Form (e.g., PTO/SB/17)  
(Submit an original and a duplicate for fee processing)
2. ☒ Specification [Total Pages **26**]  
(preferred arrangement set forth below)
  - Descriptive title of the Invention
  - Cross References to Related Applications
  - Statement Regarding Fed sponsored R & D
  - Reference to Microfiche Appendix
  - Background of the Invention
  - Brief Summary of the Invention
  - Brief Description of the Drawings (if filed)
  - Detailed Description
  - Claim(s)
  - Abstract of the Disclosure
3. ☒ Drawing(s) (35 U.S.C. 113) [Total Sheets **12**]
4. Oath or Declaration [Total Pages **3**]
  - a. ☐ Newly executed (original or copy)
  - b. ☒ Copy from a prior application (37 C.F.R. § 1.63(d))  
(for continuation/divisional with Box 16 completed)
    - i. ☐ **DELETION OF INVENTOR(S)**  
Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. §§ 1.63(d)(2) and 1.33(b).

\* NOTE FOR ITEMS 1 & 13: IN ORDER TO BE ENTITLED TO PAY SMALL ENTITY FEES, A SMALL ENTITY STATEMENT IS REQUIRED (37 C.F.R. § 1.27), EXCEPT IF ONE FILED IN A PRIOR APPLICATION IS RELIED UPON (37 C.F.R. § 1.28).

## ADDRESS TO:

Assistant Commissioner for Patents  
Box Patent Application  
Washington, DC 20231

5. ☐ Microfiche Computer Program (Appendix)
6. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
  - a. ☐ Computer Readable Copy
  - b. ☒ Paper Copy (identical to computer copy)
  - c. ☒ Statement verifying identity of above copies

## ACCOMPANYING APPLICATION PARTS

7. ☒ Assignment Papers (cover sheet & document(s))
8. ☐ 37 C.F.R. § 3.73(b) Statement ☐ Power of Attorney  
(when there is an assignee)
9. ☐ English Translation Document (if applicable)
10. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
11. ☒ Preliminary Amendment
12. ☒ Return Receipt Postcard (MPEP 503)  
(Should be specifically itemized)
13. ☐ \* Small Entity Statement(s) ☒ Statement filed in prior application, Status still proper and desired  
(PTO/SB/09-12)
14. ☐ Certified Copy of Priority Document(s)  
(if foreign priority is claimed)
15. ☒ Other: **Claim of Priority**  
**Copy of Small Entity Declaration**

## 16. If a CONTINUING APPLICATION, check appropriate box, and supply the requisite information below and in a preliminary amendment:

- ☒ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: **08 / 819,612**  
Prior application information: Examiner **S. Ungar** Group / Art Unit: **1642**

For CONTINUATION or DIVISIONAL APPS only: The entire disclosure of the prior application, from which an oath or declaration is supplied under Box 4b, is considered a part of the disclosure of the accompanying continuation or divisional application and is hereby incorporated by reference. The incorporation can only be relied upon when a portion has been inadvertently omitted from the submitted application parts.

## 17. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label

(Inse.



label here)

or ☐ Correspondence address below

Name

**021269**

Address

PATENT AND TRADEMARK OFFICE

City

State

Zip Code

Country

Telephone

Fax

Name (Print/Type)

**Gilberto M. Villacorta, Ph.D.**

Registration No. (Attorney/Agent)

**34,038**

Signature

*G. Villacorta*

Date

**07.19.00**

Burden Hour Statement. This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231

Applicant or Patentee: Ole Thastrup, et al.  
Attorney's Docket No.: 114465.400  
Serial or Patent No.: 08/819,612  
Filed or Issued: March 17, 1997  
For: NOVEL FLUORESCENT PROTEINS

**VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITY STATUS  
(37) CFR 1.9(f) and 1.27(c)) - SMALL BUSINESS CONCERN**

I hereby declare that I am

- ☐ the owner of the small business concern identified below:  
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF ORGANIZATION: BioImage A/S

ADDRESS OF ORGANIZATION: Mørkhøj Bygade 28, 2860 Søborg, Denmark

I hereby declare that the above identified small business concern qualified as a small business concern as defined in 13 CFR 121.3-18, and reproduced in 37 CFR 1.9(d), for purposes of paying reduced fees under section 41(a) and (b) of Title 35, United States Code, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby declare that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention, entitled by inventor(s) described in

- ☐ the specification filed herewith.  
☒ application Serial No. 08/819,612 filed March 17, 1997  
☐ patent no. \_\_\_\_\_, issued \_\_\_\_\_.

If the rights held by the above identified small business concern are not exclusive, each individual, concern or organization having rights to the invention is listed below\* and no rights to the invention are held by any person, other than the inventor, who could not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e). \*NOTE: Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_


☐ INDIVIDUAL      ☐ SMALL BUSINESS CONCERN      ☐ NONPROFIT ORGANIZATION

NAME \_\_\_\_\_

ADDRESS \_\_\_\_\_

☐ INDIVIDUAL      ☐ SMALL BUSINESS CONCERN      ☐ NONPROFIT ORGANIZATION

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

NAME OF PERSON SIGNING: Ole Thastrup  Jørgen Ulrik Veitsgaard  
TITLE IN ORGANIZATION: Chief Technology Officer Chief Operating Officer  
ADDRESS OF PERSON SIGNING: Mørkhøj Bygade 28, DK-2860 Søborg, Denmark  
SIGNATURE:  
DATE: 6 June 2000

[illegible]

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of: Thastrup et al.

Serial No.: 08/819,612

Group Art Unit: 1806

Filed: March 17, 1997

Examiner: Ungar, S.

For: Novel Fluorescent Proteins

**VERIFIED STATEMENT UNDER 37 CFR 1.821(f)**

Hon. Commissioner of Patents and Trademarks  
Washington, DC 20231


Sir:

I hereby state that the content of the paper and computer readable copies of the Sequence Listing, submitted in accordance with 37 CFR § 1.821(c) and (e), respectively, are the same.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issue thereon.

Respectfully submitted,

Date: October 15, 1997



Valeta A. Gregg, Reg. No. 35,127  
Novo Nordisk of North America, Inc.  
405 Lexington Avenue, Suite 6400  
New York, NY 10174-6401  
(212) 867-0123

006720-07661950

Docket No.: 114465.401

PATENT

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of :  
THASTRUP et al. :  
Serial No. : Group Art Unit:  
Filed: : Examiner:  
For: NOVEL FLUORESCENT PROTEINS :

PRELIMINARY AMENDMENT

Honorable Commissioner of  
Patents and Trademarks  
Washington, D.C. 20231

Sir:

Preliminary to examination of the above-referenced application, please amend the application as follows:

**IN THE SPECIFICATION:**

Please amend the specification as follows:

Page 1, after CROSS-REFERENCE TO RELATED APPLICATIONS, please delete "This application" and insert in its place

--This application is a continuation of U.S. Application No. 08/819,612, filed March 17, 1997, which--.

**IN THE CLAIMS:**

Please amend the claims as follows:

Claim 18, line 1, delete "2, 3, 4, 5, 6 or 7";

Claim 19, line 1, delete "or 15";

Docket No.: 114465.401

Claim 20, line 1, delete "2, 3, 4, 5, 6 or 7";

Claim 21, line 1, delete "2, 3, 4, 5, 6 or 7"; and

Claim 22, line 1, delete "2, 3, 4, 5, 6 or 7".

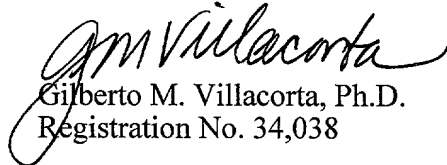
**AUTHORIZATION**

The Commissioner is hereby authorized to charge any additional fees which may be required for this Amendment, or credit any overpayment to deposit account no. 50-0436.

In the event that an extension of time is required, or which may be required in addition to that requested in a petition for an extension of time, the Commissioner is requested to grant a petition for that extension of time which is required to make this response timely and is hereby authorized to charge any fee for such an extension of time or credit any overpayment for an extension of time to deposit account no. 50-0436.

Respectfully Submitted,

PEPPER HAMILTON LLP

  
Gilberto M. Villacorta, Ph.D.  
Registration No. 34,038

Hamilton Square  
600 Fourteenth Street  
Washington, DC 20005  
202.220.1200 GMV:lrr  
Date: 07-29-00  
Facsimile: 202-220-1665  
DC: #155105 v1 (3B\_H01 WPD)

## NOVEL FLUORESCENT PROTEINS

### 5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of PCT/DK96/00051 filed January 31, 1996 and claims priority of Danish application serial no. 1065/95 filed 22 September 1995, the contents of which applications are fully incorporated herein by reference.

### 10 FIELD OF THE INVENTION

The present invention relates to novel variants of the fluorescent protein GFP having improved fluorescence properties.

### BACKGROUND OF THE INVENTION

15 The discovery that Green Fluorescent Protein (GFP) from the jellyfish *A. victoria* retains its fluorescent properties when expressed in heterologous cells has provided biological research with a new, unique and powerful tool (Chalfie et al (1994). Science 263:802; Prasher (1995) Trends in Genetics 11:320; WO 95/07463).

20 Furthermore, the discovery of a blue fluorescent variant of GFP (Heim et al. (1994). Proc.Natl.Acad.Sci. 91:12501) has greatly increased the potential applications of using fluorescent recombinant probes to monitor cellular events or functions, since the availability of probes having different excitation and emission spectra permits simultaneous monitoring of more than one process.

25 However, the blue fluorescing variant described by Heim et al, Y66H-GFP, suffers from certain limitations: The blue fluorescence is weak (emission maximum at 448nm), thus making detection difficult, and necessitating prolonged excitation of cells expressing Y66H-GFP. Moreover, the prolonged period of excitation is damaging to cells especially because the excitation wavelength is in the UV range, 360nm - 390nm.

30 A very important aspect of using recombinant, fluorescent proteins in studying cellular functions is the non-invasive nature of the assay. This allows detection of cellular events in intact, living cells. A limitation with current fluorescent proteins is, however, that relatively high intensity light sources are needed for visualization. Especially with the

blue variant, Y66H-GFP, it is necessary to excite with intensities that are damaging to most cells. It is worth mentioning that some cellular events like oscillations in intracellular signalling systems, e.g. cytosolic free calcium, are very photo sensitive. A further consequence of the low light emittance is that only high levels of expression can be detected. Obtaining such high level expression may stress the transcriptional and/or translational machinery of the cells.

The excitation spectrum of the green fluorescent protein from *Aequorea victoria* shows two peaks: A major peak at 396nm, which is in the potentially cell damaging UV range, and a lesser peak at 475nm, which is in an excitation range that is much less harmful to cells. Heim et al.(1995), Nature, Vol. 373, p. 663-4, discloses a Ser65Thr mutation of GFP (S65T) having longer wavelengths of excitation and emission, 490nm and 510nm, respectively, than the wild-type GFP and wherein the fluorophore formation proceeded about fourfold more rapidly than in the wild-type GFP.

Expression of GFP or its fluorescent variants in living cells provides a valuable tool for studying cellular events and it is well known that many cells, including mammalian cells, are incubated at approximately 37°C in order to secure optimal and/or physiologically relevant growth. Cell lines originating from different organisms or tissues may have different relevant temperatures ranging from about 35°C for fibroblasts to about 38°C - 39°C for mouse  $\beta$ -cells. Experience has shown, however, that the fluorescent signal from cells expressing GFP is weak or absent when said cells are incubated at temperatures above room temperature, cf. Webb, C.D. et al., Journal of Bacteriology, Oct. 1995, p. 5906-5911. Ogawa H. et al., Proc. Natl. Acad. Sci. USA, Vol. 92, pp. 11899-11903, December 1995, and Lim et al. J. Biochem. 118, 13-17 (1995). The improved fluorescent variant S65T described by Heim et al. (1995) supra also displays very low fluorescence when incubated under normal culture conditions (37°C), cf. Kaether and Gerdes FEBS Letters 369 (1995) pp. 267-271. Many experiments involving the study of cell metabolism are dependent on the possibility of incubating the cells at physiologically relevant temperatures, i.e. temperatures at about 37°C.

## SUMMARY OF THE INVENTION

The purpose of the present invention is to provide novel fluorescent proteins, such as F64L-GFP, F64L-Y66H-GFP and F64L-S65T-GFP that result in a cellular



fluorescence far exceeding the cellular fluorescence from cells expressing the parent proteins, i.e. GFP, the blue variant Y66H-GFP and the S65T-GFP variant, respectively. This greatly improves the usefulness of fluorescent proteins in studying cellular functions in living cells.

5 A further purpose of the invention is to provide novel fluorescent proteins that exhibit high fluorescence in cells expressing them when said cells are incubated at a temperature of 30°C or above, preferably at a temperature of from 32°C to 39°C, more preferably at a temperature of from 35°C to 38°C, and most preferably at a temperature of about 37°C.

10 It is known that fluorescence in wild-type GFP is due to the presence of a chromophore, which is generated by cyclisation and oxidation of the SYG at position 65-67 in the predicted primary amino acid sequence and presumably by the same reasoning of the SHG sequence and other GFP analogues at positions 65-67, cf. Heim et al. (1994). Surprisingly, we have found that a mutation, preferably a substitution, of the F amino acid residue at position 1 preceding the S of the SYG or SHG chromophore or the T of the THG chromophore, *in casu* position 64 in the predicted primary amino acid sequence, results in a substantial increase of fluorescence intensity apparently without shifting the excitation and emission wavelengths. This increase is remarkable for the blue variant Y66H-GFP, which hitherto has not been useful in biological systems because of its weak fluorescence.

15 20 The F64L, F64I, F64V, F64A, and F64G substitutions are preferred, the F64L substitution being most preferred, but other mutations, e.g. deletions, insertions, or posttranslational modifications immediately preceding the chromophore are also included in the invention, provided that they result in improved fluorescence properties of the various fluorescent proteins. It should be noted that extensive deletions may result in loss of the fluorescent properties of GFP. It has been shown, that only one residue can be sacrificed from the amino terminus and less than 10 or 15 from the carboxyl terminus before fluorescence is lost, cf. Cubitt et al. TIBS Vol. 20 (11), pp. 448-456, November 1995.

25 30 Accordingly, one aspect of the present invention relates to a fluorescent protein derived from *Aequorea* Green Fluorescent Protein (GFP) or any functional analogue thereof, wherein the amino acid in position 1 upstream from the chromophore has been

mutated to provide an increase of fluorescence intensity when the fluorescent protein of the invention is expressed in cells. Surprisingly, said mutation also results in a significant increase of the intensity of the fluorescent signal from cells expressing the mutated GFP and incubated at 30°C or above 30°C, preferably at about 37°C, compared to the prior art GFP variants.

There are several advantages of the proteins of the invention, including:

Excitation with low energy light sources. Due to the high degree of brightness of F64L-Y66H-GFP and F64L-GFP their emitted light can be detected even after excitation with low energy light sources. Thereby it is possible to study cellular phenomena, such as oscillations in intracellular signalling systems, that are sensitive to light induced damage. As the intensity of the emitted light from the novel blue and green emitting fluorescent proteins are of the same magnitude, it is possible to visualize them simultaneously using the same light source.

A real time reporter for gene expression in living cells is now possible, since the fluorescence from F64L-Y66H-GFP and F64L-GFP reaches a detectable level much faster than from wild type GFP, and prior known derivatives thereof. Hence, it is more suitable for real time studies of gene expression in living cells. Detectable fluorescence may be obtained faster due to shorter maturation time of the chromophore, higher emission intensity, or a more stable protein or a combination thereof.

Simultaneous expression of the novel fluorescent proteins under control of two or more separate promoters.

Expression of more than one gene can be monitored simultaneously without any damage to living cells.

Simultaneous expression of the novel proteins using one reporter as internal reference and the other as variable marker, since regulated expression of a gene can be monitored quantitatively by fusion of a promoter to e.g. F64L-GFP (or F64L-Y66H-GFP), measuring the fluorescence, and normalizing it to the fluorescence of constitutively expressed F64L-Y66H-GFP (or F64L-GFP). The constitutively expressed F64L-Y66H-GFP (or F64L-GFP) works as internal reference.

Use as a protein tag in living and fixed cells. Due to the strong fluorescence the novel proteins are suitable tags for proteins present at low concentrations. Since no

substrate is needed and visualisation of the cells do not damage the cells dynamic analysis can be performed.

Use as an organelle tag. More than one organelle can be tagged and visualised simultaneously in living cells, e.g. the endoplasmic reticulum and the cytoskeleton.

5 Use as markers in cell or organelle fusions. By labelling two or more cells or organelles with the novel proteins, e.g. F64L-Y66H-GFP and F64L-GFP, respectively, fusions, such as heterokaryon formation, can be monitored.

10 Translocation of proteins fused to the novel proteins of the invention can be visualised. The translocation of intracellular proteins to a specific organelle, can be visualised by fusing the protein of interest to one fluorescent protein, e.g. F64L-Y66H-GFP, and labelling the organelle with another fluorescent protein, e.g. F64L-GFP, which emits light of a different wavelength. Translocation can then be detected as a spectral shift of the fluorescent proteins in the specific organelle.

15 Use as a secretion marker. By fusion of the novel proteins to a signal peptide or a peptide to be secreted, secretion may be followed on-line in living cells. A precondition for that is that the maturation of a detectable number of novel fluorescent protein molecules occurs faster than the secretion. This appears not to be the case for the fluorescent proteins GFP or Y66H-GFP of the prior art.

20 Use as genetic reporter or protein tag in transgenic animals. Due to the strong fluorescence of the novel proteins, they are suitable as tags for proteins and gene expression, since the signal to noise ratio is significantly improved over the prior art proteins, such as wild-type GFP.

25 Use as a cell or organelle integrity marker. By co-expressing two of the novel proteins, the one targeted to an organelle and the other expressed in the cytosol, it is possible to calculate the relative leakage of the cytosolic protein and use that as a measure of cell integrity.

30 Use as a marker for changes in cell morphology. Expression of the novel proteins in cells allows easy detection of changes in cell morphology, e.g. blebbing, caused by cytotoxic agents or apoptosis. Such morphological changes are difficult to visualize in intact cells without the use of fluorescent probes.

Use as a transfection marker, and as a marker to be used in combination with FACS sorting. Due to the increased brightness of the novel proteins the quality of cell detection and sorting can be significantly improved.

Use of the novel proteins as a ratio real-time kinase probe. By simultaneous expression of, e.g. F64L-GFP (or F64L-Y66H-GFP), which emits more light upon phosphorylation and a derivative of F64L-Y66H-GFP which emits less light upon phosphorylation. Thereby, the ratio of the two intensities would reveal kinase activity more accurately than only one probe.

Use as real-time probe working at near physiological concentrations. Since the novel proteins are significantly brighter than wild type GFP and prior art derivatives at about 37°C the concentration needed for visualisation can be lowered. Target sites for enzymes engineered into the novel proteins, e.g. F64L-Y66H-GFP or F64L-GFP, can therefore be present in the cell at low concentrations in living cells. This is important for two reasons: 1) The probe must interfere as little as possible with the intracellular process being studied; 2) the translational and transcriptional apparatus should be stressed minimally.

The novel proteins can be used as real time probes based on energy transfer. A probe system based on energy transfer from, e.g. F64L-Y66H-GFP to F64L-GFP.

The novel proteins can be used as reporters to monitor live/dead biomass of organisms, such as fungi. By constitutive expression of F64L-Y66H-GFP or F64L-GFP in fungi the viable biomass will light up.

Transposon vector mutagenesis can be performed using the novel proteins as markers in transcriptional and translational fusions.

Transposons to be used in microorganisms encoding the novel proteins. The transposons may be constructed for translational and transcriptional fusions. To be used for screening for promoters.

Transposon vectors encoding the novel proteins, such as F64L-Y66H-GFP and F64L-GFP, can be used for tagging plasmids and chromosomes.

Use of the novel proteins enables the study of transfer of conjugative plasmids, since more than one parameter can be followed in living cells. The plasmid may be tagged by F64L-Y66H-GFP or F64L-GFP and the chromosome of the donor/recipient by F64L-Y66H-GFP or F64L-GFP.

Use as a reporter for bacterial detection by introducing the novel proteins into the genome of bacteriophages.

By engineering the novel proteins, e.g. F64L-Y66H-GFP or F64L-GFP, into the genome of a phage a diagnostic tool can be designed. F64L-Y66H-GFP or F64L-GFP will be expressed only upon transfection of the genome into a living host. The host specificity is defined by the bacteriophage.

Any novel feature or combination of features described herein is considered essential to this invention.

## DETAILED DESCRIPTION OF THE INVENTION.

In a preferred embodiment of the present invention, the novel fluorescent protein is the F64L mutant of GFP or the blue variant Y66H-GFP, said mutant showing increased fluorescence intensity. A preferred sequence of the gene encoding GFP derived from *Aequorea victoria* is disclosed in Fig. 2 herein. Fig. 2 shows the nucleotide sequence of a wild-type GFP (Hind3-EcoR1 fragment) and the amino acid sequence, wherein start codon ATG corresponds to position 8 and stop codon TAA corresponds to position 722 in the nucleotide sequence. A microorganism, *E. coli* NN049087, carrying the DNA sequence shown in Fig. 2 has been deposited for the purpose of patent procedure according to the Budapest Treaty in Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroderweg 1 b, D-38124 Braunschweig, Federal Republic of Germany, under the deposition No. DSM 10260. Another sequence of an isotype of this gene is disclosed by Prasher et al., Gene **111**, 1992, pp. 229-233 (GenBank Accession No. M62653). Besides, the novel fluorescent proteins may also be derived from other fluorescent proteins, e.g. the fluorescent protein of the sea pansy *Renilla reniformis*.

Herein the abbreviations used for the amino acids are those stated in J. Biol. Chem. 243 (1968), 3558.

The DNA construct of the invention encoding the novel fluorescent proteins may be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by Beaucage and Caruthers, Tetrahedron Letters **22** (1981), 1859 - 1869, or the method described by Matthes et al., EMBO Journal **3** (1984), 801 - 805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

The DNA construct may also be prepared by polymerase chain reaction (PCR) using specific primers, for instance as described in US 4,683,202 or Saiki et al., Science 239 (1988), 487-491. A more recent review of PCR methods may be found in PCR Protocols, 1990, Academic Press, San Diego, California, USA.

5

The DNA construct of the invention may be inserted into a recombinant vector which may be any vector which may conveniently be subjected to recombinant DNA procedures. The choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

10

The vector is preferably an expression vector in which the DNA sequence encoding the fluorescent protein of the invention is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the fluorescent protein of the invention.

15

20

The promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell, including native *Aequorea* GFP genes.

25

Examples of suitable promoters for directing the transcription of the DNA sequence encoding the fluorescent protein of the invention in mammalian cells are the SV40 promoter (Subramani et al., Mol. Cell Biol. 1 (1981), 854-864), the MT-1 (metallothionein gene) promoter (Palmiter et al., Science 222 (1983), 809-814) or the adenovirus 2 major late promoter.

30

An example of a suitable promoter for use in insect cells is the polyhedrin promoter (US 4,745,051; Vasudevan et al., FEBS Lett. 311, (1992) 7-11), the P10 promoter (J.M. Vlak et al., J. Gen. Virology 69, 1988, pp. 765-776), the *Autographa californica* polyhedrosis virus basic protein promoter (EP 397 485), the baculovirus

immediate early gene 1 promoter (US 5,155,037; US 5,162,222), or the baculovirus 39K delayed-early gene promoter (US 5,155,037; US 5,162,222).

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al., J. Biol. Chem. 255 (1980), 12073 - 12080; Alber and Kawasaki, J. Mol. Appl. Gen. 1 (1982), 419 - 434) or alcohol dehydrogenase genes (Young et al., in Genetic Engineering of Microorganisms for Chemicals (Hollaender et al, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al., Nature 304 (1983), 652 - 654) promoters.

Examples of suitable promoters for use in filamentous fungus host cells are, for instance, the ADH3 promoter (McKnight et al., The EMBO J. 4 (1985), 2093 - 2099) or the tpiA promoter. Examples of other useful promoters are those derived from the gene encoding *A. oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *A. niger* neutral  $\alpha$ -amylase, *A. niger* acid stable  $\alpha$ -amylase, *A. niger* or *A. awamori* glucoamylase (gluA), *Rhizomucor miehei* lipase, *A. oryzae* alkaline protease, *A. oryzae* triose phosphate isomerase or *A. nidulans* acetamidase. Preferred are the TAKA-amylase and gluA promoters.

Examples of suitable promoters for use in bacterial host cells include the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the *Bacillus licheniformis* alpha-amylase gene, the *Bacillus amyloliquefaciens* BAN amylase gene, the *Bacillus subtilis* alkaline protease gene, or the *Bacillus pumilus* xylosidase gene, or by the phage Lambda P<sub>R</sub> or P<sub>L</sub> promoters or the *E. coli* lac, trp or tac promoters.

The DNA sequence encoding the novel fluorescent proteins of the invention may also, if necessary, be operably connected to a suitable terminator, such as the human growth hormone terminator (Palmiter et al., op. cit.) or (for fungal hosts) the TPI1 (Alber and Kawasaki, op. cit.) or ADH3 (McKnight et al., op. cit.) terminators. The vector may further comprise elements such as polyadenylation signals (e.g. from SV40 or the adenovirus 5' Elb region), transcriptional enhancer sequences (e.g. the SV40 enhancer) and translational enhancer sequences (e.g. the ones encoding adenovirus VA RNAs).

The recombinant vector may further comprise a DNA sequence enabling the vector to replicate in the host cell in question. An example of such a sequence (when the host cell is a mammalian cell) is the SV40 origin of replication.

When the host cell is a yeast cell, suitable sequences enabling the vector to replicate are the yeast plasmid 2 $\mu$  replication genes REP 1-3 and origin of replication.

The vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the *Schizosaccharomyces pombe* TPI gene (described by P.R. Russell, Gene 40, 1985, pp. 125-130), or one which confers resistance to a drug, e.g. ampicillin, kanamycin, tetracyclin, chloramphenicol, neomycin or hygromycin. For filamentous fungi, selectable markers include amdS, pyrG, argB, niaD, sC.

The procedures used to ligate the DNA sequences coding for the fluorescent protein of the invention, the promoter and optionally the terminator and/or secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., op.cit.).

The host cell into which the DNA construct or the recombinant vector of the invention is introduced may be any cell which is capable of expressing the present DNA construct and includes bacteria, yeast, fungi and higher eukaryotic cells.

Examples of bacterial host cells which, on cultivation, are capable of expressing the DNA construct of the invention are grampositive bacteria, e.g. strains of *Bacillus*, such as *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus*, *B. megatherium* or *B. thuringiensis*, or strains of *Streptomyces*, such as *S. lividans* or *S. murinus*, or gramnegative bacteria such as *Echerichia coli*. The transformation of the bacteria may be effected by protoplast transformation or by using competent cells in a manner known per se (cf. Sambrook et al., supra).

Examples of suitable mammalian cell lines are the HEK293 and the HeLa cell lines, primary cells, and the COS (e.g. ATCC CRL 1650), BHK (e.g. ATCC CRL 1632, ATCC CCL 10), CHL (e.g. ATCC CCL39) or CHO (e.g. ATCC CCL 61) cell lines. Methods of transfecting mammalian cells and expressing DNA sequences introduced in the cells are described in e.g. Kaufman and Sharp, J. Mol. Biol. **159** (1982), 601 - 621; Southern and Berg, J. Mol. Appl. Genet. **1** (1982), 327 - 341; Loyter et al., Proc. Natl. Acad. Sci. USA **79** (1982), 422 - 426; Wigler et al., Cell **14** (1978), 725; Corsaro and



Pearson, Somatic Cell Genetics 7 (1981), 603, Graham and van der Eb, Virology 52 (1973), 456; and Neumann et al., EMBO J. 1 (1982), 841 - 845.

Examples of suitable yeast cells include cells of *Saccharomyces* spp. or *Schizosaccharomyces* spp., in particular strains of *Saccharomyces cerevisiae* or *Saccharomyces kluyveri*. Methods for transforming yeast cells with heterologous DNA and producing heterologous polypeptides therefrom are described, e.g. in US 4,599,311, US 4,931,373, US 4,870,008, 5,037,743, and US 4,845,075, all of which are hereby incorporated by reference. Transformed cells are selected by a phenotype determined by a selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient, e.g. leucine. A preferred vector for use in yeast is the POT1 vector disclosed in US 4,931,373. The DNA sequence encoding the fluorescent protein of the invention may be preceded by a signal sequence and optionally a leader sequence, e.g. as described above. Further examples of suitable yeast cells are strains of *Kluyveromyces*, such as *K. lactis*, *Hansenula*, e.g. *H. polymorpha*, or *Pichia*, e.g. *P. pastoris* (cf. Gleeson et al., J. Gen. Microbiol. 132, 1986, pp. 3459-3465; US 4,882,279).

Examples of other fungal cells are cells of filamentous fungi, e.g. *Aspergillus* spp., *Neurospora* spp., *Fusarium* spp. or *Trichoderma* spp., in particular strains of *A. oryzae*, *A. nidulans* or *A. niger*. The use of *Aspergillus* spp. for the expression of proteins is described in, e.g., EP 272 277, EP 230 023, EP 184 438.

When a filamentous fungus is used as the host cell, it may be transformed with the DNA construct of the invention, conveniently by integrating the DNA construct in the host chromosome to obtain a recombinant host cell. This integration is generally considered to be an advantage as the DNA sequence is more likely to be stably maintained in the cell. Integration of the DNA constructs into the host chromosome may be performed according to conventional methods, e.g. by homologous or heterologous recombination.

Transformation of insect cells and production of heterologous polypeptides therein may be performed as described in US 4,745,051; US 4,879,236; US 5,155,037; 5,162,222; EP 397,485) all of which are incorporated herein by reference. The insect cell line used as the host may suitably be a *Lepidoptera* cell line, such as *Spodoptera frugiperda* cells or *Trichoplusia ni* cells (cf. US 5,077,214). Culture conditions may

suitably be as described in, for instance, WO 89/01029 or WO 89/01028, or any of the aforementioned references.

The transformed or transfected host cell described above is then cultured in a suitable nutrient medium under conditions permitting the expression of the present DNA construct after which the cells may be used in the screening method of the invention. Alternatively, the cells may be disrupted after which cell extracts and/or supernatants may be analysed for fluorescence.

The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. in catalogues of the American Type Culture Collection).

In the method of the invention, the fluorescence of cells transformed or transfected with the DNA construct of the invention may suitably be measured in a spectrometer or a fluorescence microscope where the spectral properties of the cells in liquid culture may be determined as scans of light excitation and emission.

The invention is further illustrated in the following examples with reference to the appended drawings.

Example 1.

#### Cloning of cDNA encoding GFP

Briefly, total RNA, isolated from *A. victoria* by a standard procedure (Sambrook et al., Molecular Cloning. 2., eds. (1989) (Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York), 7.19-7.22) was converted into cDNA by using the AMV reverse transcriptase (Promega, Madison, WI, USA) as recommended by the manufacturer. The cDNA was then PCR amplified, using PCR primers designed on the basis of a previously published GFP sequence (Prasher et al., **Gene** **111** (1992), 229-223; GenBank accession No. M62653) together with the UITma™ polymerase (Perkin Elmer, Foster City, CA, USA). The sequences of the primers were: GFP2:  
TGGAATAAGCTTTATGAGTAAAGGAGAAGAAGCTTTT and GFP-1:  
AAGAATTCGGATCCCTTTAGTGTCAATTGGAAGTCT  
Restriction endonuclease sites inserted in the 5' (a HindIII site) and 3' (EcoRI and BamHI sites) primers facilitated the cloning of the PCR amplified GFP cDNA into a

slightly modified pUC19 vector. The details of the construction are as follows: LacZ Shine-Dalgarno AGGA, immediately followed by the 5' HindIII site plus an extra T and the GFP ATG codon, giving the following DNA sequence at the lacZ-promoter GFP fusion point:  $P_{LacZ}$ -AGGAAAGCTTTATG-GFP. At the 3' end of the GFP cDNA, the  
5 base pair corresponding to nucleotide 770 in the published GFP sequence (GenBank accession No. M62653) was fused to the EcoRI site of the pUC19 multiple cloning site (MCS) through a PCR generated BamHI, EcoRI linker region).

The DNA sequence and predicted primary amino acid sequence of GFP is shown below in Fig. 2a. Another DNA sequence encoding the same amino acid sequence as  
10 shown in Fig. 2a is shown in Fig. 2b. To generate the blue fluorescent variant described by Heim et al. (1994), a PCR primer incorporating the Y66H substitution responsible for changing green fluorescence into blue fluorescence was used as 5' PCR primer in combination with a GFP specific 3' primer. The template was the GFP clone described above. The sequence of the 5' primer is 5'-

15 CTACCTGTTCCATGGCCAACGCTTGTCACCTTTTCCTCATGGTGTTCAATGCTT  
TTCTAGATACCC-3' (SEQ ID NO:3). Its 5' end corresponds to position 164 in the GFP sequence. In addition to the Y66H substitution, the 5' primer introduces a A to T change at position 223 ; this mutation creates a XbaI site without changing an amino acid. The 5' primer also contains the naturally occurring NcoI recognition sequence  
20 (position 173 in the GFP sequence). The sequence of the 3' primer is 5'-

AAGAATTCGGATCCCTTTAGTGTCAATTGGAAGTCT-3' (SEQ ID NO:4). Position  
3 from the 5' end is the first base of the EcoRI recognition site that corresponds to the 3' end of the GFP sequence. The resulting PCR product was digested with NcoI and EcoRI and cloned into an NcoI-EcoRI vector fragment to reconstitute the entire Y66H-  
25 GFP gene.

E.coli cells carrying an expression vector containing Y66H-GFP were grown overnight in the presence of 10 micrograms per ml N-methyl-N-nitro-N-nitrosoguanidine. Plasmid DNA was isolated, the 764 bp Hind3-EcoRI insert containing Y66H-GFP was isolated and cloned into a Hind3-EcoRI digested vector fragment, allowing expression of  
30 the insert in E.coli. E.coli transformants were inspected for blue fluorescence when excited with a 365 nm UV light, and colonies that appeared to fluoresce stronger than wildtype BFP were identified.

10 ng DNA from one particular colony was used as template in a PCR reaction containing 1.5 units of Taq polymerase (Perkin Elmer), 0.1mM MnCl<sub>2</sub>, 0.2 mM each of dGTP, dCTP and dTTP, 0.05mM dATP, 1.7 mM MgCl<sub>2</sub> and the buffer recommended by the manufacturer. The primers used flank the Y66H-GFP insert. The sequence of the 5' primer was 5'-AATTGGTACCAAGGAGGTAAGCTTTATGAG-3' (SEQ ID NO:5); it contains a Hind3 recognition sequence. The sequence of the 3' primer was 5'-CTTTCGTTTTGAATTCGGATCCCTTTAGTG-3' (SEQ ID NO:6); it contains a EcoR1 recognition sequence.

The PCR product was digested with Hind3 and EcoR1 and cloned into a Hind3-EcoR1 digested vector fragment, allowing expression of the insert in E.coli. E.coli transformants were inspected for blue fluorescence when excited with a 365 nm UV light, and colonies that appeared to fluoresce stronger than Y66H-GFP were identified. Plasmid DNA from one strongly fluorescing colony (called BX12-1A) was isolated and the Y66H-GFP insert was subjected to sequence determination. The mutation F64L was identified. This mutation replaces the phenylalanine residue preceding the SHG tripeptide chromophore sequence of Y66H-GFP with leucine. No other amino acid changes were present in the Y66H-GFP sequence of BX12-1A. The DNA sequence and predicted primary amino acid sequence of F64L-Y66H-GFP is shown in Fig. 3 below.

## Example 2.

F64L-GFP was constructed as follows: An E.coli expression vector containing Y66H-GFP was digested with restriction enzymes Nco1 and Xba1. The recognition sequence of Nco1 is located at position 173 and the recognition sequence of Xba1 is located at position 221 in the F64L-Y66H-GFP sequence listed below. The large Nco1-Xba1 vector fragment was isolated and ligated with a synthetic Nco1-Xba1 DNA linker of the following sequence:

One DNA strand has the sequence:

5'-CATGGCCAACGCTTGTCACTACTCTCTTATGGTGTTCAATGCTTTT-3' (SEQ ID NO:7)

The other DNA strand has the sequence:

5'-CTAGAAAAGCATTGAACACCATAAGAGAGAGTAGTGACAAGCGTTGGC-3' (SEQ ID NO:8)

Upon annealing, the two strands form a Nco1-Xba1 fragment that incorporates the sequence of the GFP chromophore SYG with the F64L substitution preceding SYG. The DNA sequence and predicted primary amino acid sequence of F64L-GFP is shown in Fig. 4 below.

The S65T-GFP mutation was described by Heim et al (Nature vol.373 pp. 663-664, 1995). F64L-S65T-GFP was constructed as follows: An E.coli expression vector containing Y66H-GFP was digested with restriction enzymes Nco1 and Xba1. The recognition sequence of Nco1 is located at position 173 and the recognition sequence of Xba1 is located at position 221 in the F64L-Y66H-GFP sequence listed below. The large Nco1-Xba1 vector fragment was isolated and ligated with a synthetic Nco1-Xba1 DNA linker of the following sequence:

One DNA strand has the sequence:

5'-CATGGCCAACGCTTGTCACCTACTCTCACTTATGGTGTTCAATGCTTTT-3' (SEQ ID NO:9)

The other DNA strand has the sequence:

5'-CTAGAAAAGCATTGAACACCATAAGTGAGAGTAGTGACAAGCGTTGGC-3' (SEQ ID NO:10).

Upon annealing, the two strands form a Nco1-Xba1 fragment that incorporates the F64L and S65T mutations in the GFP chromophore. The DNA sequence and predicted primary amino acid sequence of F64L-S65T-GFP is shown in Fig. 5 below.

The E. coli expression vector contains an IPTG (isopropyl-thio-galactoside)-inducible promoter. The E. coli strain used is a del(lacZ)MI5 derivative of K 803 (Sambrook et al. *supra*).

The GFP allele present in the pGFP-N1 plasmid (available from Clontech Laboratories) was introduced into the IPTG inducible E.coli expression vector in the following manner:

1 ng pGFP-N1 plasmid DNA was used as template in a standard PCR reaction where the 5' PCR primer had the sequence:

5'- TGGAATAAGCTTTATGAGTAAAGGAGAAGAACTTTT - 3' (SEQ ID NO:11)

and the 3' PCR primer had the sequence:

5' - GAATCGTAGATCTTTATTTGTATAGTTCATCCATG - 3' (SEQ ID NO:12).

The primers flank the GFP-N1 insert in the vector pGFP-N1. The 5' primer includes the ATG start codon preceded by a Hind3 cloning site. The 3' primer includes a TAA stop codon followed by a Bgl2 cloning site.

The PCR product was digested with Hind3 and Bgl2 and cloned into a Hind3-BamH1 digested vector fragment behind an IPTG inducible promoter, allowing expression of the insert in E.coli in the presence of IPTG.

The lacZ gene present in the pZeoSV-LacZ plasmid (available from Invitrogen) was introduced into the IPTG inducible E.coli expression vector in the following manner:

1 ng pZeoSV-LacZ plasmid DNA was used as template in a standard PCR reaction where the 5' PCR primer had the sequence:

5' - TGGAATAAGCTTTATGGATCCCGTCGTTTACAACGTCGT - 3' (SEQ ID NO:13)

and the 3' PCR primer had the sequence:

5' - GCGCGAATTCTTATTATTATTTTACACACCAGAC - 3' (SEQ ID NO:14).

The primers flank the lacZ insert in the vector pZeoSV-LacZ. The 5' primer includes the ATG start codon preceded by a Hind3 cloning site. The 3' primer includes a TAA stop codon followed by an EcoR1 cloning site.

The PCR product was digested with Hind3 and EcoR1 and cloned into a Hind3-EcoR1 digested vector fragment behind an IPTG inducible promoter, allowing expression of the insert in E.coli in the presence of IPTG.

To measure and compare the fluorescence generated in E. coli cells expressing GFP, GFP-N1, F64L-GFP, F64L-S65T-GFP, Y66H-GFP, F64L-Y66H-GFP or beta-galactosidase (as background control) under various conditions the following experiments were done:

E. coli cells containing an expression plasmid allowing expression of one of the various gene products upon induction with IPTG were grown in LB medium containing 100 micrograms per milliliter ampicillin and no IPTG. To 1 ml cell suspension was added 0.5 ml 50% glycerol and cells were frozen and kept frozen at -80C.

Cells from the - 80C glycerol stocks were inoculated into 2 ml LB medium containing 100 µg/ml ampicillin and grown with aeration at 37C for 6 hours. 2 microliters of this inoculum was transferred to each of two tubes containing 2 ml of LB

medium with 100  $\mu\text{g/ml}$  ampicillin and 1 mM IPTG. The two sets of tubes were incubated with aeration at two different temperatures: room temperature (22C) and 37C.

After 16 hours 0.2 ml samples were taken of cells expressing GFP, GFP-N1, F64L-GFP, F64L-S65T-GFP, Y66H-GFP, F64L-Y66H-GFP or beta-galactosidase. Cells were pelleted, the supernatant was removed, cells were resuspended in 2 ml water and transferred to a cuvette. Fluorescence emission spectra were measured in a LS-50 luminometer (Perkin-Elmer) with excitation and emission slits set to 10 nm. The excitation wavelengths were set to 398 nm and 470 nm for GFP, GFP-N1, F64L-GFP and F64L-S65T-GFP; 398 nm is near the optimal excitation wavelength for GFP, GFP-N1 and F64L-GFP, and 470 nm is near the optimal excitation wavelength for F64L-S65T-GFP. For Y66H-GFP and F64L-Y66H-GFP the excitation wavelength was set to 380 nm, which is near the optimal excitation wavelength for these derivatives. Beta-galactosidase expressing cells were included as background controls. Following the measurements in the LS-50 luminometer, the optical density at 450 nm was measured for each sample in a spectrophotometer (Lambda UV/VIS, Perkin-Elmer). This is a measure of total cells in the assay. Luminometer data were normalized to the optical density of the sample.

The results of the experiments are shown in Fig. 6a - 6f below and can be summarized as follows:

After 16 hours at 22C using an excitation wavelength of 398 nm there were large signals for GFP and F64L-GFP, and detectable signals for GFP-N1 and F64L-S65T-GFP, cf. Fig. 6a.

After 16 hours at 37C with an excitation wavelength of 398 nm there was a large signals for F64L-GFP, a detectable signal for F64L-S65T-GFP, and no detectable signals for GFP and GFP-N1, cf. Fig. 6b.

After 16 hours at 22C with an excitation wavelength of 470 nm there was a large signals for F64L-S65T-GFP, detectable signals for GFP and F64L-GFP, and no detectable signals for GFP-N1, cf. Fig. 6c.

After 16 hours at 37C with an excitation wavelength of 470 nm there were large signals for F64L-S65T-GFP and F64L-GFP, and no detectable signals for GFP and GFP-N1, cf. Fig. 6d.

After 16 hours at 22C with an excitation wavelength of 380 nm there were detectable signals over background for Y66H-GFP and F64L-Y66H-GFP, cf. Fig. 6e.

After 16 hours at 37C with an excitation wavelength of 380 nm there was no detectable signal over background for Y66H-GFP and a large signal for F64L-Y66H-GFP, cf. Fig. 6f.

To determine whether the differences in fluorescence signals were due to differences in expression levels, total protein from the E.coli cells (0.5 OD<sub>450</sub> units) analyzed as described above was fractionated by SDS-polyacrylamide gel electrophoresis (12% Tris-glycine gels from BIO-RAD Laboratories) followed by Western blot analysis (ECL Western blotting from Amersham International) with polyclonal GFP antibodies (from rabbit). The result showed that expression levels of GFP, GFP-N1, F64L-GFP, F64L-S65T-GFP, Y66H-GFP and F64L-Y66H-GFP were identical, both at 22C and 37C. The differences in fluorescence signals are therefore not due to different expression levels.

**Example 3. Influence of the F64L substitution on GFP and its derivatives when expressed in mammalian cells.**

F64L-Y66H-GFP, F64L-GFP, and F64L-S65T-GFP were cloned into pcDNA3 (Invitrogen, Ca, USA) so that the expression was under control of the CMV promoter. Wild-type GFP was expressed from the pGFP-N1 plasmid (Clontech, Ca, USA) in which the CMV promoter controls the expression. Plasmid DNA to be used for transfection were purified using Jetstar Plasmid kit (Genomed Inc. NC, USA) and was dissolved in distilled water.

The precipitate used for the transfections were made by mixing the following components: 2 µg DNA in 44 µl of water were mixed with 50 µl 2xHBS buffer (280 mM NaCl, 1.5 mM Na<sub>2</sub>HPO<sub>4</sub>, 12 mM dextrose, 50 mM HEPES) and 6.2 µl 2M CaCl<sub>2</sub>. The transfection mix was incubated at room temperature for 25 minutes before it was added to the cells. HEK 293 cells (ATCC CRL 1573) were grown in 2 cm by 2 cm coverglass chambers (Nunc, Denmark) with approximately 1.5 ml medium (Dulbecco's MEM with glutamax-1, 4500 mg/L glucose, and 10% FCS; Gibco BRL, MD, USA). The DNA was added to cells at 25-50% confluence. Cells were grown at 37°C in a CO<sub>2</sub> incubator. Prior to visualisation the medium was removed and 1.5 ml Ca<sup>2+</sup>-HEPES buffer (5 mM KCl,



140 mM NaCl, 5.5 mM glucose, 1 mM MgSO<sub>4</sub>, 1 mM CaCl, 10 mM HEPES) was added to the chamber.

Transfectants were visualised using an Axiovert 135 (Carl Zeiss, Germany) fluorescence microscope. The microscope was equipped with an HBO 100 mercury excitation source and a 40x, Fluar, NA = 1.3 objective (Carl Zeiss, Germany). To visualise GFP, F64L-GFP, and F64L-S65T-GFP the following filters were used: excitation 480/40 nm, dichroic 505 nm, and emission 510LP nm (all from Chroma Technologies Corp., Vt, USA). To visualise F64L-Y66H-GFP the following filters were used: excitation 380/15 nm, dichroic 400 nm, and emission 450/65 nm (all from Omega Optical, Vt, USA).

Cells in several chambers were transfected in parallel, so that, a new chamber could be taken for each sample point. In cases where the incubation extended beyond 8.5 hours the Ca<sup>2+</sup> precipitate was removed by replacing the medium.

As shown in Table 1 the F64L mutation enhances the fluorescent signal significantly (wild type GFP versus F64L-GFP and F64L-S65T-GFP). Fluorescent cells can be observed as early as 1-2 hours post-transfection indicating an efficient maturation of the chromophore at 37°C. Furthermore, the F64L mutation is enhancing other GFP derivatives like the S65T mutant which has a shifted excitation spectrum and the blue derivative which is not detectable in mammalian cells without the F64L substitution. (Comment: When comparing the results of F64L-S65T-GFP and F64L-GFP one has to take into account that the excitation spectra differ and that the filter set used is optimised for F64L-S65T-GFP. F64L-GFP and WT GFP share the same spectral properties.)

# SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Thastrup, Ole  
Tullin, Søren  
Poulsen, Lars Kongsbak  
Bjørn, Sara Petersen
- (ii) TITLE OF INVENTION: Novel Fluorescent Proteins
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Novo Nordisk of North America, Inc.
  - (B) STREET: 405 Lexington Avenue, Suite 6400
  - (C) CITY: New York
  - (D) STATE: New York
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 10174-6401
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: to be assigned
  - (B) FILING DATE: 17-MAR-1997
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Gregg, Valeta A.
  - (B) REGISTRATION NUMBER: 35,127
  - (C) REFERENCE/DOCKET NUMBER: 4594.204-US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 212-867-0123
  - (B) TELEFAX: 212-878-9655

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGGAATAAGC TTTATGAGTA AAGGAGAAGA ACTTTT

36

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAGAATTTCGG ATCCCTTTAG TGTCAATTGG AAGTCT

36

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTACCTGTTC CATGGCCAAC GCTTGTCAC TTTCTCTCA TGGTGTTC AA TGCTTTTCTA  
GATACCC

60  
67

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGAATTCGG ATCCCTTTAG TGTCAATTGG AAGTCT

36

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AATTGGTACC AAGGAGGTAA GCTTTATGAG

30

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTTTCGTTTT GAATTCGGAT CCCTTTAGTG

30

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CATGGCCAAC GCTTGTCAC TTTCTCTTT ATGGTGTTC AATGCTTTT

48

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTAGAAAAGC ATTGAACACC ATAAGAGAGA GTAGTGACAA GCGTTGGC

48

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CATGGCCAAC GCTTGTCAC TACTCTACTT ATGGTGTTC AATGCTTTT

48

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAGAAAAGC ATTGAACACC ATAAGTGAGA GTAGTGACAA GCGTTGGC

48

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGGAATAAGC TTTATGAGTA AAGGAGAAGA ACTTTT

36

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAATCGTAGA TCTTTATTTG TATAGTTCAT CCATG

35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGAATAAGC TTTATGGATC CCGTCGTTTT ACAACGTCGT

40

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 nucleotides
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

5 GCGCGAATTC TTATTATTAT TTTTGACACC AGAC

34

00640340.074000

## Claims

1. A fluorescent protein derived from Green Fluorescent Protein (GFP) or any functional analogue thereof, wherein the amino acid in position 1 preceding the chromophore has been mutated to provide an increase in fluorescence intensity.
2. A fluorescent protein according to claim 1, wherein the chromophore is in position 65-67 of the predicted primary amino acid sequence of GFP.
3. A fluorescent protein according to claim 1 resulting in an increased fluorescence in cells expressing said fluorescent protein when said cells are incubated at a temperature of 30°C or above 30°C, preferably at a temperature of from 32°C to 39°C, more preferably at a temperature of from 35°C to 38°C, and most preferably at a temperature of about 37°C.
4. A fluorescent protein according to claim 1, said protein being derived from *Aequorea victoria* or *Renilla reniformis*.
5. A fluorescent protein according to claim 1, wherein the amino acid F in position 64 of GFP or Y66H-GFP has been substituted by an amino acid selected from the group consisting of L, I, V, A and G.
6. A fluorescent protein according to claim 1, wherein the amino acid F in position 1 preceding the chromophore has been substituted by L and the amino acids of the chromophore include SYG, SHG or TYG.
7. A fluorescent protein according to claim 1 and having the amino acid sequence of Fig. 3, Fig. 4 or Fig. 5 herein.
8. A fusion compound consisting of a fluorescent protein (GFP) according to claim 1, wherein said GFP is linked to a polypeptide.
9. A fusion compound according to claim 8 wherein the polypeptide is a kinase, preferably the catalytic subunit of protein kinase A, or protein kinase C, or Erk1, or a cytoskeletal element.
10. A nucleotide sequence coding for the Fluorescent Protein of claim 1.
11. A nucleotide sequence according to claim 10 selected from the sequences shown in Fig. 3, Fig. 4 and Fig. 5.
12. A DNA construct comprising a suitable control region or regions and a nucleotide sequence according to claim 10, the sequence being under the control of the control region.
13. A DNA construct according to claim 12 being under the control of the native GFP promoter, or a mammal constitutive or regulatory promoter, a viral promoter, a yeast promoter, a filamentous fungi promoter, or a bacterial promoter.

14. A host transformed with a DNA construct according to claim 12.

15. A host according to claim 14 selected from the following: organisms and cells belonging to bacteria, yeast, fungi, protozoans and higher eucaryots .

16. A process for preparing a polypeptide, comprising cultivating a host according to claim 14 and obtaining therefrom the polypeptide expressed by said nucleotide sequence.

17. A process according to claim 16 wherein the expression of the nucleotide sequence is effected by the native GFP promoter.

18. Use of the Fluorescent Protein of claim 1, 2, 3, 4, 5, 6 or 7 in an *in vitro* assay for measuring protein kinase activity, or dephosphorylation activity, wherein said fluorescent protein in purified form is added to a biological sample, preferably a cell extract, and any change in fluorescence is recorded.

19. Use of the host of claim 14 or 15 in an *in vivo* assay for measuring metabolic activity, preferably kinase activity and dephosphorylating activity.

20. Use of the fluorescent protein of claim 1, 2, 3, 4, 5, 6 or 7 as a reporter for gene expression in living cells.

21. Use of the fluorescent protein of claim 1, 2, 3, 4, 5, 6 or 7 for the simultaneous monitoring of more than one gene in living, intact cells.

22. Use of two or more of the fluorescent protein of claim 1, 2, 3, 4, 5, 6 or 7 as organelle or cell tags for simultaneous visualisation of organelle or cell processes in living cells.

# Abstract

The present invention relates to novel variants of the fluorescent protein GFP having improved fluorescence properties.



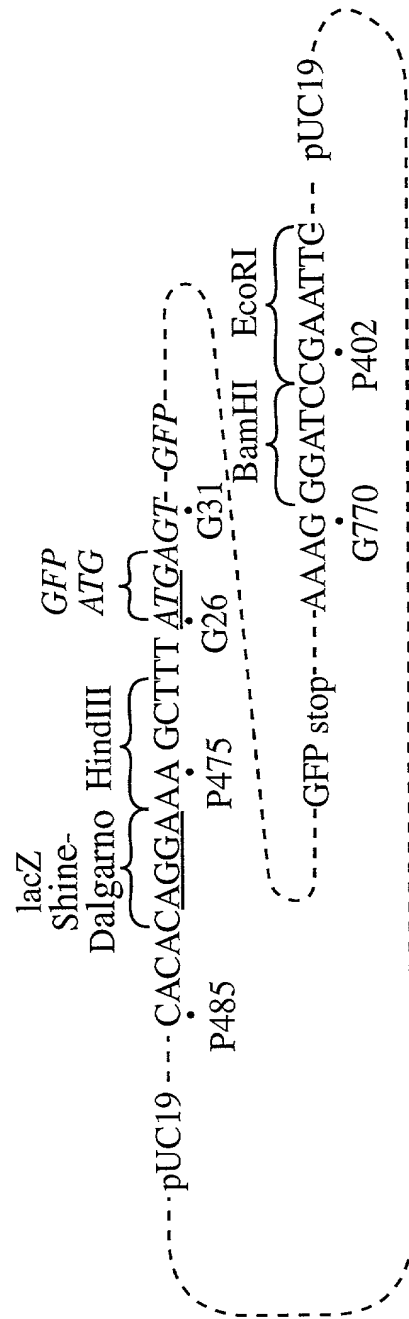


FIG. 1

## DNA and Predicted primary amino acid sequence of GFP (Hind3-EcoR1 fragment).

5' - AAGCTTT

ATG	AGT	AAA	GGA	GAA	GAA	CTT	TTC	ACT	GGA	GTT	GTC	CCA	ATT	CTT	GTT	GAA	TTA	GAT	CGC
MET	SER	LYS	GLY	GLU	GLU	LEU	PHE	THR	GLY	VAL	VAL	PRO	ILE	LEU	VAL	GLU	LEU	ASP	GLY
GAT	GTT	AAT	GGG	CAA	AAA	TTC	TCT	GTT	AGT	GGA	GAG	GGT	GAA	GGT	GAT	GCA	ACA	TAC	GGA
ASP	VAL	ASN	GLY	GLN	LYS	PHE	SER	VAL	SER	GLY	GLU	GLY	GLU	GLY	ASP	ALA	THR	TYR	GLY
AAA	CTT	ACC	CTT	AAA	TTT	ATT	TGC	ACT	ACT	GGG	AAG	CTA	CCT	GTT	CCA	TGG	CCA	ACG	CTT
LYS	LEU	THR	LEU	LYS	PHE	ILE	CYS	THR	THR	GLY	LYS	LEU	PRO	VAL	PRO	TRP	PRO	THR	LEU
GTC	ACT	ACT	TTC	TCT	TAT	GGT	GTT	CAA	TGC	TTT	TCA	AGA	TAC	CCA	GAT	CAT	ATG	AAA	CAG
VAL	THR	THR	PHE	SER	TYR	GLY	VAL	GLN	CYS	PHE	SER	ARG	TYR	PRO	ASP	HIS	MET	LYS	GLN
CAT	GAC	TTT	TTC	AAG	AGT	GCC	ATG	CCC	GAA	GGT	TAT	GTA	CAG	GAA	AGA	ACT	ATA	TTT	TAC
HIS	ASP	PHE	PHE	LYS	SER	ALA	MET	PRO	GLU	GLY	TYR	VAL	GLN	GLU	ARG	THR	ILE	PHE	TYR
AAA	GAT	GAC	GGG	AAC	TAC	AAG	ACA	CGT	GCT	GAA	GTC	AAG	TTT	GAA	GGT	GAT	ACC	CTT	GTT
LYS	ASP	ASP	GLY	ASN	TYR	LYS	THR	ARG	ALA	GLU	VAL	LYS	PHE	GLU	GLY	ASP	THR	LEU	VAL
AAT	AGA	ATC	GAG	TTA	AAA	GGT	ATT	GAT	TTT	AAA	GAA	GAT	GGA	AAC	ATT	CTT	GGA	CAC	AAA
ASN	ARG	ILE	GLU	LEU	LYS	GLY	ILE	ASP	PHE	LYS	GLU	ASP	GLY	ASN	ILE	LEU	GLY	HIS	LYS
ATG	GAA	TAC	AAT	TAT	AAC	TCA	CAC	AAT	GTA	TAC	ATC	ATG	GCA	GAC	AAA	CCA	AAG	AAT	GGA
MET	GLU	TYR	ASN	TYR	ASN	SER	HIS	ASN	VAL	TYR	ILE	MET	ALA	ASP	LYS	PRO	LYS	ASN	GLY
ATC	AAA	GTT	AAC	TTC	AAA	ATT	AGA	CAC	AAC	ATT	AAA	GAT	GGA	AGC	GTT	CAA	TTA	GCA	GAC
ILE	LYS	VAL	ASN	PHE	LYS	ILE	ARG	HIS	ASN	ILE	LYS	ASP	GLY	SER	VAL	GLN	LEU	ALA	ASP
CAT	TAT	CAA	CAA	AAT	ACT	CCA	ATT	GGC	GAT	GGC	CCT	GTC	CTT	TTA	CCA	GAC	AAC	CAT	TAC
HIS	TYR	GLN	GLN	ASN	THR	PRO	ILE	GLY	ASP	GLY	PRO	VAL	LEU	LEU	PRO	ASP	ASN	HIS	TYR
CTG	TCC	ACG	CAA	TCT	GCC	CTT	TCC	AAA	GAT	CCC	AAC	GAA	AAG	AGA	GAT	CAC	ATG	ATC	CTT
LEU	SER	THR	GLN	SER	ALA	LEU	SER	LYS	ASP	PRO	ASN	GLU	LYS	ARG	ASP	HIS	MET	ILE	LEU
CTT	GAG	TTT	GTA	ACA	GCT	GCT	GGG	ATT	ACA	CAT	GGC	ATG	GAT	GAA	CTA	TAC	AAA	TAA	
LEU	GLU	PHE	VAL	THR	ALA	ALA	GLY	ILE	THR	HIS	GLY	MET	ASP	GLU	LEU	TYR	LYS		

ATGTCCAGACTTCCAATTGACACTAAAGGGATCCGAATTC - 3'

Fig. 2a

Nucleotide sequence (764bp) of GFP (Hind3-EcoR1 fragment)

AAGCTTTATGAGTAAAGGAGAAGAACTTTTCACTGGAGTT  
GTCCAATTCTTGTTGAATTAGATGGCGATGTTAATGGGC  
AAAAATTCTCTGTTAGTGGAGAGGGTGAAGGTGATGCAAC  
ATACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGG  
AAGCTACCTGTTCCATGGCCAACGCTTGTCCTACTTTCT  
CTTATGGTGTTCATGCTTTTCAAGATACCCAGATCATAT  
GAAACAGCATGACTTTTTCAAGAGTGCCATGCCCCGAAGGT  
TATGTACAGGAAAGAAGTATATTTTACAAAGATGACGGGA  
ACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATAC  
CCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAA  
GAAGATGGAAACATTCTTGGACACAAAATGGAATACAAC  
ATAACTCACATAATGTATACATCATGGCAGACAAACCAA  
GAATGGCATCAAAGTTAACTTCAAAATTAGACACAACATT  
AAAGATGGAAGCGTTCAATTAGCAGACCATTATCAACAAA  
ATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAA  
CCATTACCTGTCCACGCAATCTGCCCTTCCAAAGATCCC  
AACGAAAAGAGAGATCATATGATCCTTCTTGAGTTTGTA  
CAGCTGCTGGGATTACACATGGCATGGATGAACTATACAA  
ATAAATGTCCAGACTTCCAATTGACACTAAAGGGATCCGA  
ATTC

Fig. 2b

## DNA and predicted primary amino acid sequence of F64L-Y66H-GFP (Hind3-EcoR1 fragment).

5' - AAGCTTT

ATG	AGT	AAA	GGA	GAA	GAA	CTT	TTC	ACT	GGA	GTT	GTC	CCA	ATT	CTT	GTT	GAA	TTA	GAT	GGC
MET	SER	LYS	GLY	GLU	GLU	LEU	PHE	THR	GLY	VAL	VAL	PRO	ILE	LEU	VAL	GLU	LEU	ASP	GLY
GAT	GTT	AAT	GGG	CAA	AAA	TTC	TCC	GTT	AGT	GGA	GAG	GGT	GAA	GGT	GAT	GCA	ACA	TAC	GGA
ASP	VAL	ASN	GLY	GLN	LYS	PHE	SER	VAL	SER	GLY	GLU	GLY	GLU	GLY	ASP	ALA	THR	TYR	GLY
AAA	CTT	ACC	CTT	AAA	TTT	ATT	TGC	ACT	ACT	GGG	AAG	CTA	CCT	GTT	CCA	TGG	CCA	ACG	CTT
LYS	LEU	THR	LEU	LYS	PHE	ILE	CYS	THR	THR	GLY	LYS	LEU	PRO	VAL	PRO	TRP	PRO	THR	LEU
GTC	ACT	ACT	CTC	TCT	CAT	GGT	GTT	CAA	TGC	TTT	TCT	AGA	TAC	CCA	GAT	CAT	ATG	AAA	CAG
VAL	THR	THR	LEU	SER	HIS	GLY	VAL	GLN	CYS	PHE	SER	ARG	TYR	PRO	ASP	HIS	MET	LYS	GLN
CAT	GAC	TTT	TTC	AAG	AGT	GCC	ATG	CCC	GAA	GGT	TAT	GTA	CAG	GAA	AGA	ACT	ATA	TTT	TAC
HIS	ASP	PHE	PHE	LYS	SER	ALA	MET	PRO	GLU	GLY	TYR	VAL	GLN	GLU	ARG	THR	ILE	PHE	TYR
AAA	GAT	GAC	GGG	AAC	TAC	AAG	ACA	CGT	GCT	GAA	GTC	AAG	TTT	GAA	GGT	GAT	ACC	CTT	GTT
LYS	ASP	ASP	GLY	ASN	TYR	LYS	THR	ARG	ALA	GLU	VAL	LYS	PHE	GLU	GLY	ASP	THR	LEU	VAL
AAT	AGA	ATC	GAG	TTA	AAA	GGT	ATT	GAT	TTT	AAA	GAA	GAT	GGA	AAC	ATT	CTT	GGA	CAC	AAA
ASN	ARG	ILE	GLU	LEU	LYS	GLY	ILE	ASP	PHE	LYS	GLU	ASP	GLY	ASN	ILE	LEU	GLY	HIS	LYS
ATG	GAA	TAC	AAT	TAT	AAC	TCA	CAT	AAT	GTA	TAC	ATC	ATG	GCA	GAC	AAA	CCA	AAG	AAT	GGC
MET	GLU	TYR	ASN	TYR	ASN	SER	HIS	ASN	VAL	TYR	ILE	MET	ALA	ASP	LYS	PRO	LYS	ASN	GLY
ATC	AAA	GTT	AAC	TTC	AAA	ATT	AGA	CAC	AAC	ATT	AAA	GAT	GGA	AGC	GTT	CAA	TTA	GCA	GAC
ILE	LYS	VAL	ASN	PHE	LYS	ILE	ARG	HIS	ASN	ILE	LYS	ASP	GLY	SER	VAL	GLN	LEU	ALA	ASP
CAT	TAT	CAA	CAA	AAT	ACT	CCA	ATT	GGC	GAT	GGC	CCT	GTC	CTT	TTA	CCA	GAC	AAC	CAT	TAC
HIS	TYR	GLN	GLN	ASN	THR	PRO	ILE	GLY	ASP	GLY	PRO	VAL	LEU	LEU	PRO	ASP	ASN	HIS	TYR
CTG	TCC	ACG	CAA	TCT	GCC	CTT	TCC	AAA	GAT	CCC	AAC	GAA	AAG	AGA	GAT	CAC	ATG	ATC	CTT
LEU	SER	THR	GLN	SER	ALA	LEU	SER	LYS	ASP	PRO	ASN	GLU	LYS	ARG	ASP	HIS	MET	ILE	LEU
CTT	GAG	TTT	GTA	ACA	GCT	GCT	GGG	ATT	ACA	CAT	GGC	ATG	GAT	GAA	CTA	TAC	AAA	TAA	
LEU	GLU	PHE	VAL	THR	ALA	ALA	GLY	ILE	THR	HIS	GLY	MET	ASP	GLU	LEU	TYR	LYS		

ATGTCAGACTTCCAATTGACACTAAAGGGATCCGAATTC - 3'

Fig. 3

## DNA and predicted primary amino acid sequence of F64L-GFP (Hind3 – EcoR1 fragment).

5' - AAGCTTT

ATG MET	AGT SER	AAA LYS	GGA GLY	GAA GLU	GAA GLU	CTT LEU	TTC PHE	ACT THR	GGA GLY	GTT VAL	GTC VAL	CCA PRO	ATT ILE	CTT LEU	GTT VAL	GAA GLU	TTA LEU	GAT ASP	GGC GLY
GAT ASP	GTT VAL	AAT ASN	GGG GLY	CAA GLN	AAA LYS	TTC PHE	TCT SER	GTT VAL	AGT SER	GGA GLY	GAG GLU	GGT GLY	GAA GLU	GGT GLY	GAT ASP	GCA ALA	ACA THR	TAC TYR	GGA GLY
AAA LYS	CTT LEU	ACC THR	CTT LEU	AAA LYS	TTT PHE	ATT ILE	TGC CYS	ACT THR	ACT THR	GGG GLY	AAG LYS	CTA LEU	CCT PRO	GTT VAL	CCA PRO	TGG TRP	CCA PRO	ACG THR	CTT LEU
GTC VAL	ACT THR	ACT THR	CTC LEU	TCT SER	TAT TYR	GGT GLY	GTT VAL	CAA GLN	TGC CYS	TTT PHE	TCT SER	AGA ARG	TAC TYR	CCA PRO	GAT ASP	CAT HIS	ATG MET	AAA LYS	CAG GLN
CAT HIS	GAC ASP	TTT PHE	TTC PHE	AAG LYS	AGT SER	GCC ALA	ATG MET	CCC PRO	GAA GLU	GGT GLY	TAT TYR	GTA VAL	CAG GLN	GAA GLU	AGA ARG	ACT THR	ATA ILE	TTT PHE	TAC TYR
AAA LYS	GAT ASP	GAC ASP	GGG GLY	AAC ASN	TAC TYR	AAG LYS	ACA THR	CGT ARG	GCT ALA	GAA GLU	GTC VAL	AAG LYS	TTT PHE	GAA GLU	GGT GLY	GAT ASP	ACC THR	CTT LEU	GTT VAL
AAT ASN	AGA ARG	ATC ILE	GAG GLU	TTA LEU	AAA LYS	GGT GLY	ATT ILE	GAT ASP	TTT PHE	AAA LYS	GAA GLU	GAT ASP	GGA GLY	AAC ASN	ATT ILE	CTT LEU	GGA GLY	CAC HIS	AAA LYS
ATG MET	GAA GLU	TAC TYR	AAT ASN	TAT TYR	AAC ASN	TCA SER	CAT HIS	AAT ASN	GTA VAL	TAC TYR	ATC ILE	ATG MET	GCA ALA	GAC ASP	AAA LYS	CCA PRO	AAG LYS	AAT ASN	GGC GLY
ATC ILE	AAA LYS	GTT VAL	AAC ASN	TTC PHE	AAA LYS	ATT ILE	AGA ARG	CAC HIS	AAC ASN	ATT ILE	AAA LYS	GAT ASP	GGA GLY	AGC SER	GTT VAL	CAA GLN	TTA LEU	GCA ALA	GAC ASP
CAT HIS	TAT TYR	CAA GLN	CAA GLN	AAT ASN	ACT THR	CCA PRO	ATT ILE	GGC GLY	GAT ASP	GGC GLY	CCT PRO	GTC VAL	CTT LEU	TTA LEU	CCA PRO	GAC ASP	AAC ASN	CAT HIS	TAC TYR
CTG LEU	TCC SER	ACG THR	CAA GLN	TCT SER	GCC ALA	CTT LEU	TCC SER	AAA LYS	GAT ASP	CCC PRO	AAC ASN	GAA GLU	AAG LYS	AGA ARG	GAT ASP	CAC HIS	ATG MET	ATC ILE	CTT LEU
CTT LEU	GAG GLU	TTT PHE	GTA VAL	ACA THR	GCT ALA	GCT ALA	GGG GLY	ATT ILE	ACA THR	CAT HIS	GGC GLY	ATG MET	GAT ASP	GAA GLU	CTA LEU	TAC TYR	AAA LYS	TAA	

ATGTCAGACTTCCAATTGACACTAAAGGGATCCGAATTC - 3'

Fig. 4

DNA and predicted primary amino acid sequence of F64L-S65T-GFP (Hind3 – EcoR1 fragment).

5' - AAGCTTT

ATG	AGT	AAA	GGA	GAA	GAA	CTT	TTC	ACT	GGA	GTT	GTC	CCA	ATT	CTT	GTT	GAA	TTA	GAT	GGC
MET	SER	LYS	GLY	GLU	GLU	LEU	PHE	THR	GLY	VAL	VAL	PRO	ILE	LEU	VAL	GLU	LEU	ASP	GLY
GAT	GTT	AAT	GGG	CAA	AAA	TTC	TCT	GTT	AGT	GGA	GAG	GGT	GAA	GGT	GAT	GCA	ACA	TAC	GGA
ASP	VAL	ASN	GLY	GLN	LYS	PHE	SER	VAL	SER	GLY	GLU	GLY	GLU	GLY	ASP	ALA	THR	TYR	GLY
AAA	CTT	ACC	CTT	AAA	TTT	ATT	TGC	ACT	ACT	GGG	AAG	CTA	CCT	GTT	CCA	TGG	CCA	ACG	CTT
LYS	LEU	THR	LEU	LYS	PHE	ILE	CYS	THR	THR	GLY	LYS	LEU	PRO	VAL	PRO	TRP	PRO	THR	LEU
GTC	ACT	ACT	CTC	ACT	TAT	GGT	GTT	CAA	TGC	TTT	TCT	AGA	TAC	CCA	GAT	CAT	ATG	AAA	CAG
VAL	THR	THR	LEU	THR	TYR	GLY	VAL	GLN	CYS	PHE	SER	ARG	TYR	PRO	ASP	HIS	MET	LYS	GLN
CAT	GAC	TTT	TTC	AAG	AGT	GCC	ATG	CCC	GAA	GGT	TAT	GTA	CAG	GAA	AGA	ACT	ATA	TTT	TAC
HIS	ASP	PHE	PHE	LYS	SER	ALA	MET	PRO	GLU	GLY	TYR	VAL	GLN	GLU	ARG	THR	ILE	PHE	TYR
AAA	GAT	GAC	GGG	AAC	TAC	AAG	ACA	CGT	GCT	GAA	GTC	AAG	TTT	GAA	GGT	GAT	ACC	CTT	GTT
LYS	ASP	ASP	GLY	ASN	TYR	LYS	THR	ARG	ALA	GLU	VAL	LYS	PHE	GLU	GLY	ASP	THR	LEU	VAL
AAT	AGA	ATC	GAG	TTA	AAA	GGT	ATT	GAT	TTT	AAA	GAA	GAT	GGA	AAC	ATT	CTT	GGA	CAC	AAA
ASN	ARG	ILE	GLU	LEU	LYS	GLY	ILE	ASP	PHE	LYS	GLU	ASP	GLY	ASN	ILE	LEU	GLY	HIS	LYS
ATG	GAA	TAC	AAT	TAT	AAC	TCA	CAT	AAT	GTA	TAC	ATC	ATG	GCA	GAC	AAA	CCA	AAG	AAT	GGC
MET	GLU	TYR	ASN	TYR	ASN	SER	HIS	ASN	VAL	TYR	ILE	MET	ALA	ASP	LYS	PRO	LYS	ASN	GLY
ATC	AAA	GTT	AAC	TTC	AAA	ATT	AGA	CAC	AAC	ATT	AAA	GAT	GGA	AGC	GTT	CAA	TTA	GCA	GAC
ILE	LYS	VAL	ASN	PHE	LYS	ILE	ARG	HIS	ASN	ILE	LYS	ASP	GLY	SER	VAL	GLN	LEU	ALA	ASP
CAT	TAT	CAA	CAA	AAT	ACT	CCA	ATT	GGC	GAT	GGC	CCT	GTC	CTT	TTA	CCA	GAC	AAC	CAT	TAC
HIS	TYR	GLN	GLN	ASN	THR	PRO	ILE	GLY	ASP	GLY	PRO	VAL	LEU	LEU	PRO	ASP	ASN	HIS	TYR
CTG	TCC	ACG	CAA	TCT	GCC	CTT	TCC	AAA	GAT	CCC	AAC	GAA	AAG	AGA	GAT	CAC	ATG	ATC	CTT
LEU	SER	THR	GLN	SER	ALA	LEU	SER	LYS	ASP	PRO	ASN	GLU	LYS	ARG	ASP	HIS	MET	ILE	LEU
CTT	GAG	TTT	GTA	ACA	GCT	GCT	GGG	ATT	ACA	CAT	GGC	ATG	GAT	GAA	CTA	TAC	AAA	TAA	
LEU	GLU	PHE	VAL	THR	ALA	ALA	GLY	ILE	THR	HIS	GLY	MET	ASP	GLU	LEU	TYR	LYS		

ATGTCCAGACTTCCAATTGACACTAAAGGGATCCGAATTC – 3'

Fig. 5

00649340-024900

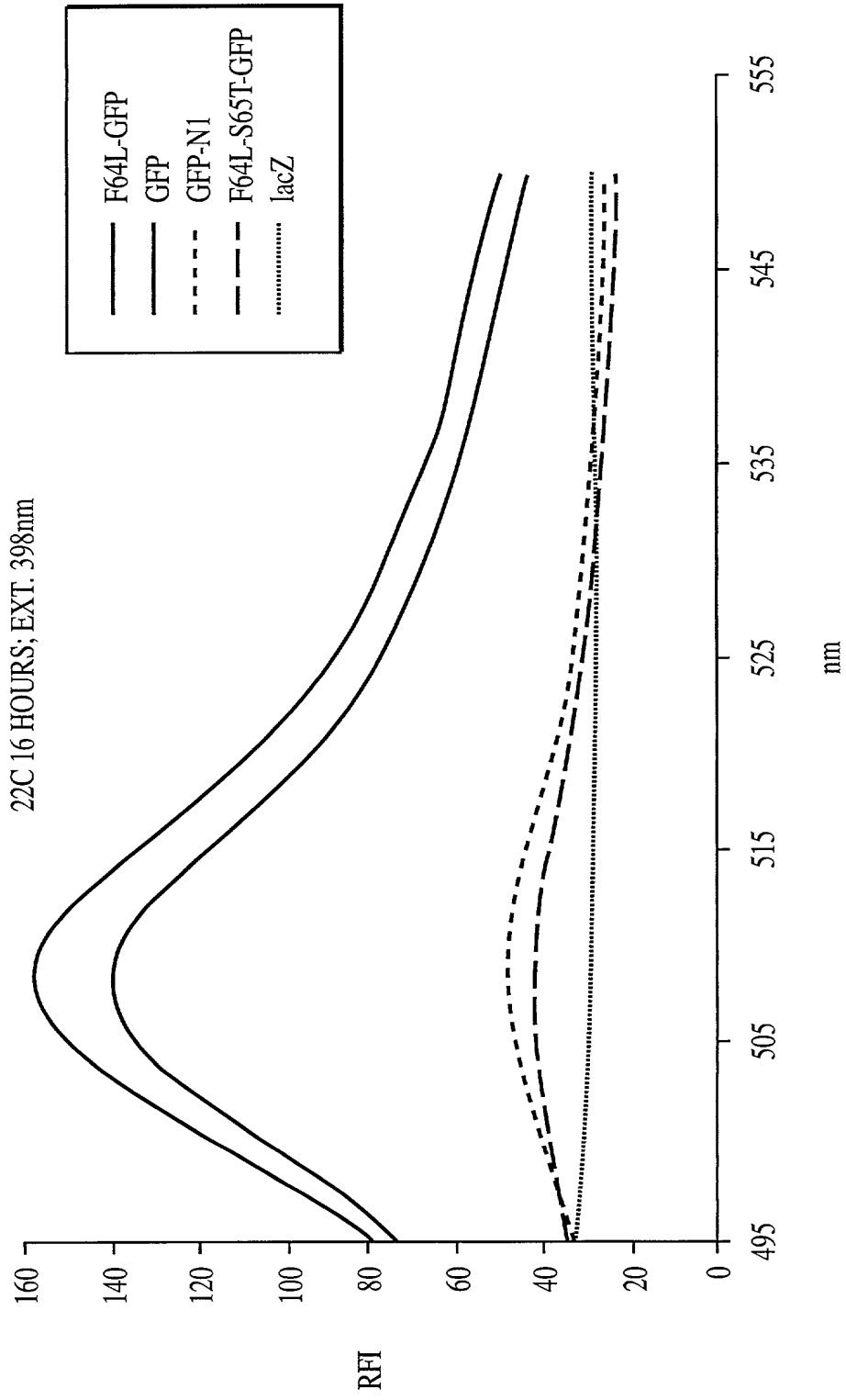


FIG. 6A

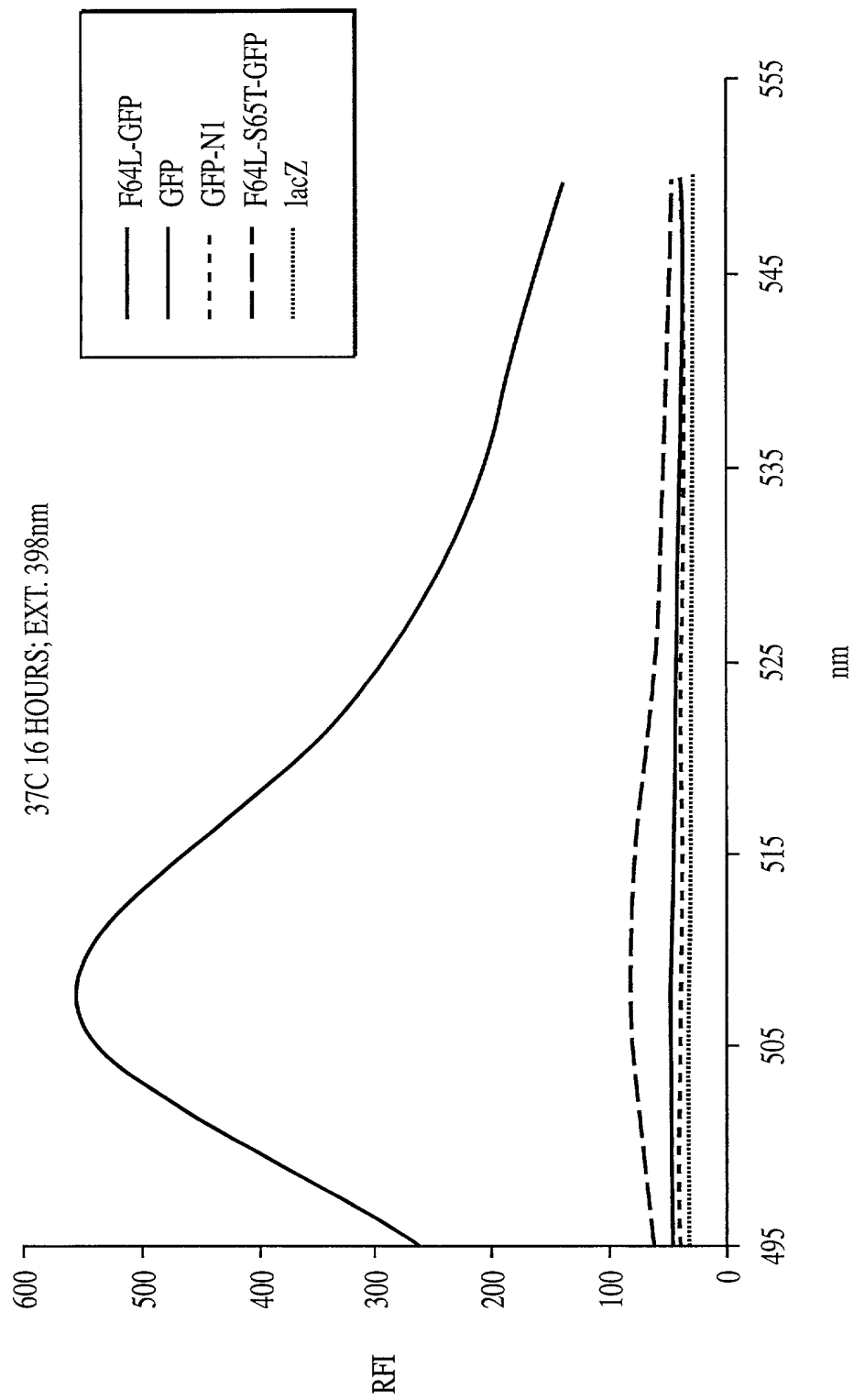


FIG. 6B



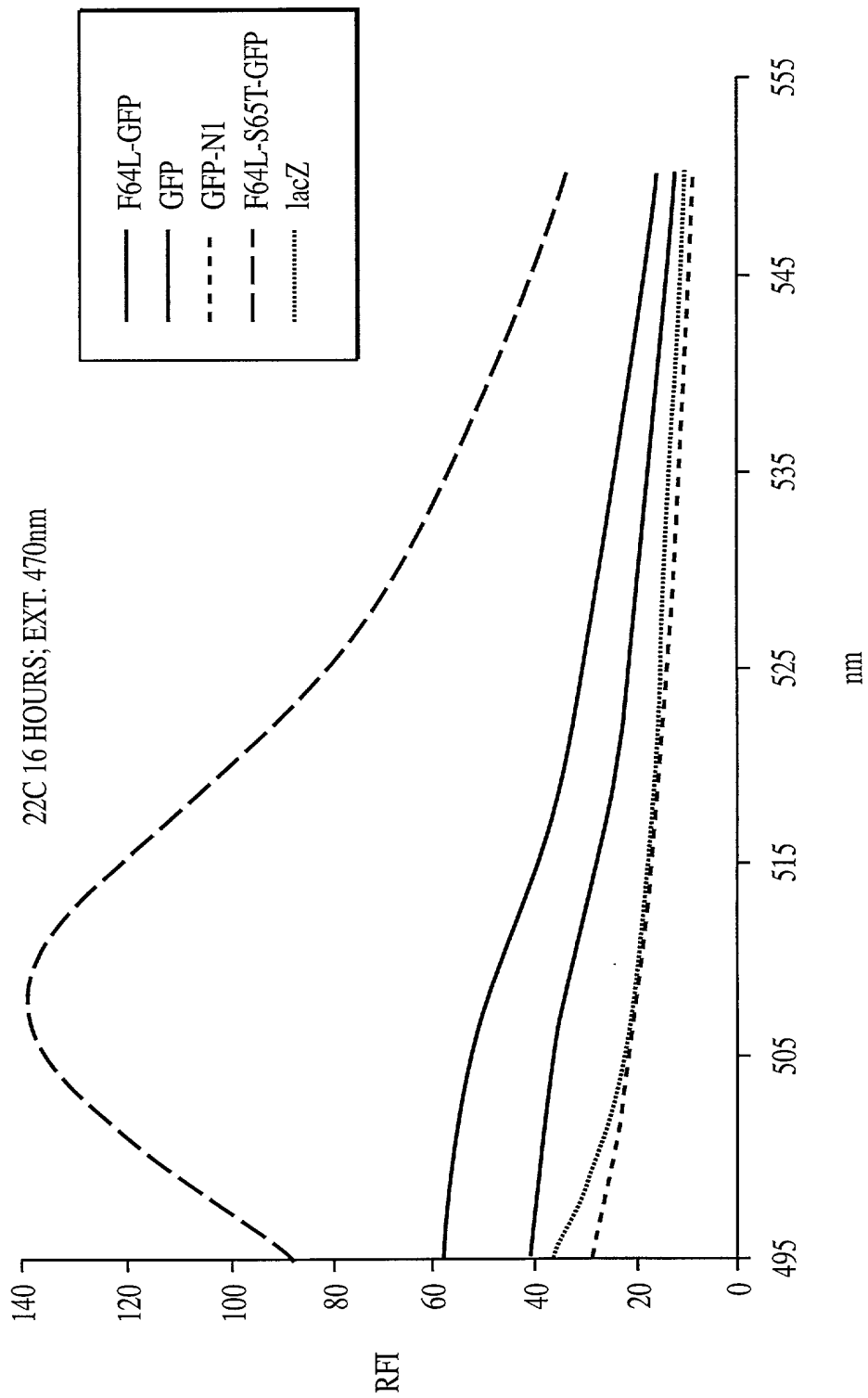


FIG. 6C

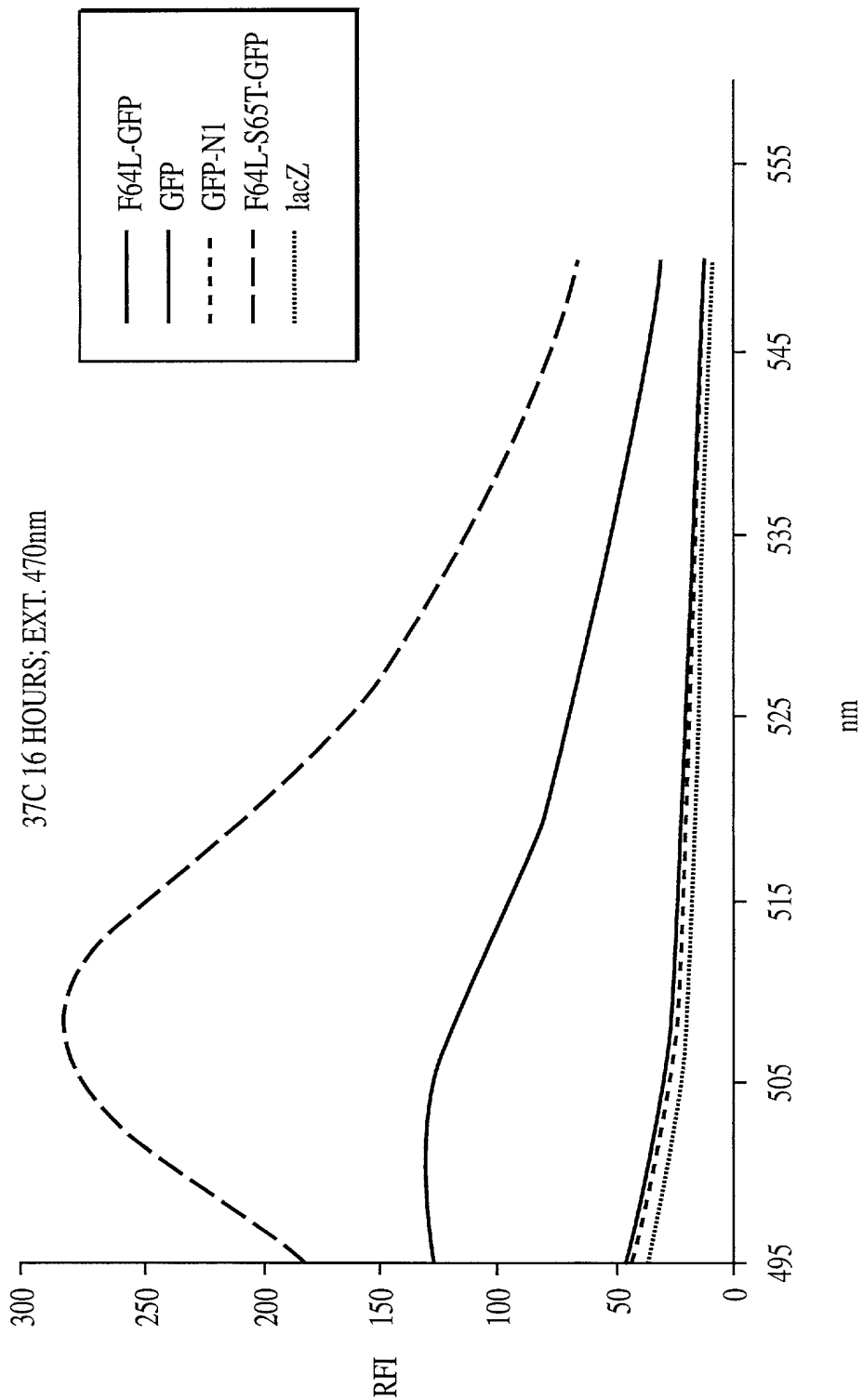


FIG. 6D

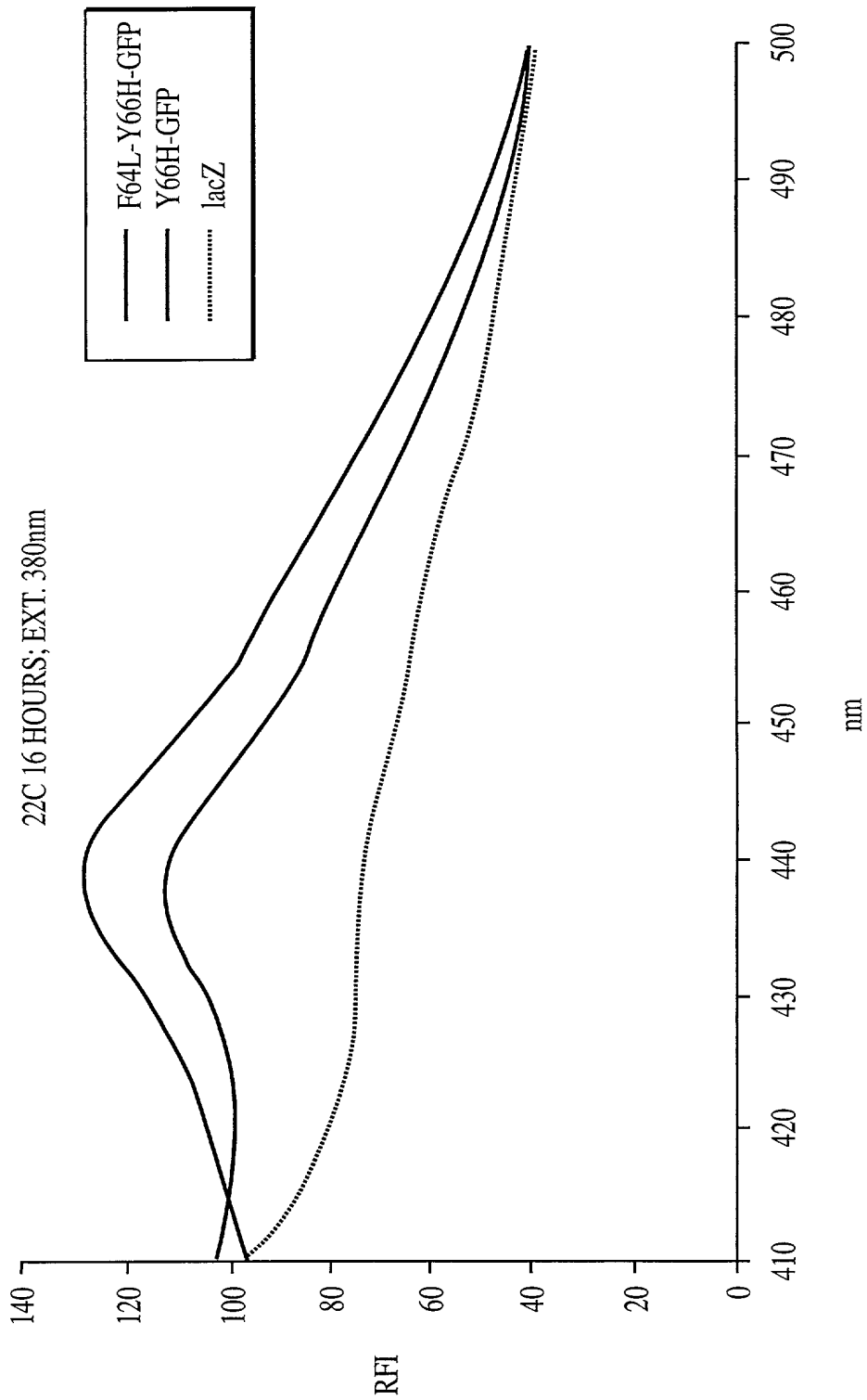


FIG. 6E



FIG. 6F

My residence, post office address and citizenship are as stated below next to my name.

## Novel Fluorescent Proteins

[ ] is attached hereto

[X] was filed as United States application

Serial No. to be assigned

on March 17, 1997

and was amended

on

[ ] was filed as PCT international application

Number

on

and was amended under PCT Article 19

он

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign applications(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

COUNTRY (if PCT, indicate "PCT")	APPLICATION NUMBER	DATE OF FILING (day, month, year)	PRIORITY CLAIMED UNDER 35 USC 119
Denmark	1065/95	22 September 1995	[x] YES [ ] NO
			[ ] YES [ ] NO
			[ ] YES [ ] NO
			[ ] YES [ ] NO
			[ ] YES [ ] NO
			[ ] YES [ ] NO

Attorney's Docket Number  
4594-204-US

PRIOR U.S. APPLICATIONS OR PCT INTERNATIONAL APPLICATIONS DESIGNATING THE U.S. FOR BENEFIT  
UNDER 35 U.S.C. 120:

APPLICATION NO.	FILING DATE	US SERIAL NUMBERS ASSIGNED (if any)
-----------------	-------------	--

Steve T. Zelson	Elias J. Lambiris	Cheryl H. Agris	James J. Harrington	Valeta A. Gregg
Req. No. 30,335	Req. No. 33,728	Req. No. 34,086	Req. No. 38,711	Req. No. 35,127

Steve T. Zelson  
(212) 867-0123

	Full Name of Inventor	Family Name	First Given Name	Second Given Name
1	Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
	Post Office Address	Post Office Address	City	State & Zip Code/Country
2	Full Name of Inventor	Family Name	First Given Name	Second Given Name
	Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
	Post Office Address	Post Office Address	City	State & Zip Code/Country
3	Full Name of Inventor	Family Name	First Given Name	Second Given Name
	Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
	Post Office Address	Post Office Address	City	State & Zip Code/Country
4	Full Name of Inventor	Family Name	First Given Name	Second Given Name
	Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
	Post Office Address	Post Office Address	City	State & Zip Code/Country

Attorney's Docket Number:  
4594-204-US

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1 <i>Christina T. Up</i>	Signature of Inventor 2 <i>Erin Tullin</i>	Signature of Inventor 3 <i>Lars Kongstakke Poulsen</i>
Date <i>970521</i>	Date <i>970521</i>	Date <i>970521</i>
Signature of Inventor 4 <i>Sara P. Bjarni</i>	Signature of Inventor 5	Signature of Inventor 6
Date <i>970521</i>	Date	Date
Signature of Inventor 7	Signature of Inventor 8	Signature of Inventor 9
Date	Date	Date

[illegible]

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Thastrup, Ole  
Tullin, Søren  
Poulsen, Lars Kongsbak  
Bjørn, Sara Petersen
- (ii) TITLE OF INVENTION: Novel Fluorescent Proteins
- (iii) NUMBER OF SEQUENCES: 20
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Novo Nordisk of North America, Inc.
  - (B) STREET: 405 Lexington Avenue, Suite 6400
  - (C) CITY: New York
  - (D) STATE: New York
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 10174-6401
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/819,612
  - (B) FILING DATE: 17-MAR-1997
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Gregg, Valeta A.
  - (B) REGISTRATION NUMBER: 35,127
  - (C) REFERENCE/DOCKET NUMBER: 4594.204-US
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 212-867-0123
  - (B) TELEFAX: 212-878-9655

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGGAATAAGC TTTATGAGTA AAGGAGAAGA ACTTTT

36

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAGAATTCGG ATCCCTTTAG TGTCAATTGG AAGTCT

36

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 67 nucleotides



	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)	(45)	(46)	(47)	(48)	(49)	(50)	(51)	(52)	(53)	(54)	(55)	(56)	(57)	(58)	(59)	(60)	(61)	(62)	(63)	(64)	(65)	(66)	(67)	(68)	(69)	(70)	(71)	(72)	(73)	(74)	(75)	(76)	(77)	(78)	(79)	(80)	(81)	(82)	(83)	(84)	(85)	(86)	(87)	(88)	(89)	(90)	(91)	(92)	(93)	(94)	(95)	(96)	(97)	(98)	(99)	(100)
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81																			

60  
67

36

30

30

48

48

- (2) INFORMATION FOR SEQ ID NO:9:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 48 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CATGGCCAAC GCTTGTCACT ACTCTCACTT ATGGTGTTC A TGCTTTT

48

- (2) INFORMATION FOR SEQ ID NO:10:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 48 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAGAAAAGC ATTGAACACC ATAAGTGAGA GTAGTGACAA GCGTTGGC

48

- (2) INFORMATION FOR SEQ ID NO:11:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 36 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGGAATAAGC TTTATGAGTA AAGGAGAAGA ACTTTT

36

- (2) INFORMATION FOR SEQ ID NO:12:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 35 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAATCGTAGA TCTTTATTG TATAGTTCAT CCATG

35

- (2) INFORMATION FOR SEQ ID NO:13:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 40 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGAATAAGC TTTATGGATC CCGTCGTTTT ACAACGTCGT

40

- (2) INFORMATION FOR SEQ ID NO:14:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 34 nucleotides  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: cDNA  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 764 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAGCTTT	ATG	AGT	AAA	GGA	GAA	GAA	CTT	TTC	ACT	GGA	GTT	GTC	CCA	ATT	CTT	52	
Met	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	Leu			
1				5					10					15			
GTT	GAA	TTA	GAT	GGC	GAT	GTT	AAT	GGG	CAA	AAA	TTC	TCC	GTT	AGT	GGA	GAG	103
Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	Gln	Lys	Phe	Ser	Val	Ser	Gly	Glu	
				20				25					30				
GGT	GAA	GGT	GAT	GCA	ACA	TAC	GGA	AAA	CTT	ACC	CTT	AAA	TTT	ATT	TGC	ACT	154
Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys	Leu	Thr	Leu	Lys	Phe	Ile	Cys	Thr	
			35				40					45					
ACT	GGG	AAG	CTA	CCT	GTT	CCA	TGG	CCA	ACG	CTT	GTC	ACT	ACT	CTC	TCT	CAT	205
Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp	Pro	Thr	Leu	Val	Thr	Thr	Leu	Ser	His	
50				55			60							65			
GGT	GTT	CAA	TGC	TTT	TCT	AGA	TAC	CCA	GAT	CAT	ATG	AAA	CAG	CAT	GAC	TTT	256
Gly	Val	Gln	Cys	Phe	Ser	Arg	Tyr	Pro	Asp	His	Met	Lys	Gln	His	Asp	Phe	
			70				75						80				
TTC	AAG	AGT	GCC	ATG	CCC	GAA	GGT	TAT	GTA	CAG	GAA	AGA	ACT	ATA	TTT	TAC	307
Phe	Lys	Ser	Ala	Met	Pro	Glu	Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Tyr	
	85				90					95						100	
AAA	GAT	GAC	GGG	AAC	TAC	AAG	ACA	CGT	GCT	GAA	GTC	AAG	TTT	GAA	GGT	GAT	358
Lys	Asp	Asp	Gly	Asn	Tyr	Lys	Thr	Arg	Ala	Glu	Val	Lys	Phe	Glu	Gly	Asp	
				105				110						115			
ACC	CTT	GTT	AAT	AGA	ATC	GAG	TTA	AAA	GGT	ATT	GAT	TTT	AAA	GAA	GAT	GGA	409
Thr	Leu	Val	Asn	Arg	Ile	Glu	Leu	Lys	Gly	Ile	Asp	Phe	Lys	Glu	Asp	Gly	
	120					125						130					
AAC	ATT	CTT	GGA	CAC	AAA	ATG	GAA	TAC	AAT	TAT	AAC	TCA	CAT	AAT	GTA	TAC	460
Asn	Ile	Leu	Gly	His	Lys	Met	Glu	Tyr	Asn	Tyr	Asn	Ser	His	Asn	Val	Tyr	
135				140					145					150			
ATC	ATG	GCA	GAC	AAA	CCA	AAG	AAT	GGC	ATC	AAA	GTT	AAC	TTC	AAA	ATT	AGA	511
Ile	Met	Ala	Asp	Lys	Pro	Lys	Asn	Gly	Ile	Lys	Val	Asn	Phe	Lys	Ile	Arg	
			155					160					165				
CAC	AAC	ATT	AAA	GAT	GGA	AGC	GTT	CAA	TTA	GCA	GAC	CAT	TAT	CAA	CAA	AAT	562
His	Asn	Ile	Lys	Asp	Gly	Ser	Val	Gln	Leu	Ala	Asp	His	Tyr	Gln	Gln	Asn	
	170				175						180					185	
ACT	CCA	ATT	GGC	GAT	GGC	CCT	GTC	CTT	TTA	CCA	GAC	AAC	CAT	TAC	CTG	TCC	613
Thr	Pro	Ile	Gly	Asp	Gly	Pro	Val	Leu	Leu	Pro	Asp	Asn	His	Tyr	Leu	Ser	
				190				195						200			
ACG	CAA	TCT	GCC	CTT	TCC	AAA	GAT	CCC	AAC	GAA	AAG	AGA	GAT	CAC	ATG	ATC	664
Thr	Gln	Ser	Ala	Leu	Ser	Lys	Asp	Pro	Asn	Glu	Lys	Arg	Asp	His	Met	Ile	
	205					210						215					
CTT	CTT	GAG	TTT	GTA	ACA	GCT	GCT	GGG	ATT	ACA	CAT	GGC	ATG	GAT	GAA	CTA	715
Leu	Leu	Glu	Phe	Val	Thr	Ala	Ala	Gly	Ile	Thr	His	Gly	Met	Asp	Glu	Leu	
220				225				230						235			
TAC	AAA	TAA	ATG	TCC	GAC	TTCCA	ATTGA	CACTAA	AAGGG	ATCCGA	ATTC						764

000120:01201950

Tyr Lys

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 238 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val  
5 10 15  
Glu Leu Asp Gly Asp Val Asn Gly Gln Lys Phe Ser Val Ser Gly Glu  
20 25 30  
Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys  
35 40 45  
Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu  
50 55 60  
Ser His Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln  
65 70 75 80  
His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg  
85 90 95  
Thr Ile Phe Tyr Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val  
100 105 110  
Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile  
115 120 125  
Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Met Glu Tyr Asn  
130 135 140  
Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Pro Lys Asn Gly  
145 150 155 160  
Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Lys Asp Gly Ser Val  
165 170 175  
Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro  
180 185 190  
Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser  
195 200 205  
Lys Asp Pro Asn Glu Lys Arg Asp His Met Ile Leu Leu Glu Phe Val  
210 215 220  
Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys  
225 230 235

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 764 nucleotides
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AAGCTTT ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC CCA ATT  
Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile  
1 5 10

CTT GTT GAA TTA GAT GGC GAT GTT AAT GGG CAA AAA TTC TCT GTT AGT	97
Leu Val Glu Leu Asp Gly Asp Val Asn Gly Gln Lys Phe Ser Val Ser	
15 20 25 30	
GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC CTT AAA TTT	145
Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe	
35 40 45	
ATT TGC ACT ACT GGG AAG CTA CCT GTT CCA TGG CCA ACG CTT GTC ACT	193
Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr	
50 55 60	
ACT CTC TCT TAT GGT GTT CAA TGC TTT TCT AGA TAC CCA GAT CAT ATG	241
Thr Leu Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met	
65 70 75	
AAA CAG CAT GAC TTT TTC AAG AGT GCC ATG CCC GAA GGT TAT GTA CAG	289
Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln	
80 85 90	
GAA AGA ACT ATA TTT TAC AAA GAT GAC GGG AAC TAC AAG ACA CGT GCT	337
Glu Arg Thr Ile Phe Tyr Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala	
95 100 105 110	
GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG TTA AAA	385
Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys	
115 120 125	
GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA ATG GAA	433
Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Met Glu	
130 135 140	
TAC AAT TAT AAC TCA CAT AAT GTA TAC ATC ATG GCA GAC AAA CCA AAG	481
Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Pro Lys	
145 150 155	
AAT GGC ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT AAA GAT GGA	529
Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Lys Asp Gly	
160 165 170	
AGC GTT CAA TTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT	577
Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp	
175 180 185 190	
GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACG CAA TCT GCC	625
Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala	
195 200 205	
CTT TCC AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC CTT CTT GAG	673
Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Ile Leu Leu Glu	
210 215 220	
TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA	721
Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys	
225 230 235	
TAA ATGTCCAGAC TTCCAATTGA CACTAAAGGG ATCCGAATTC	764

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 238 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile

5

10

006740-074000

Leu Val Glu Leu Asp Gly Asp Val Asn Gly Gln Lys Phe Ser Val Ser  
 15 20 25 30  
 Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe  
 35 40 45  
 Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr  
 50 55 60  
 Thr Leu Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met  
 65 70 75  
 Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln  
 80 85 90  
 Glu Arg Thr Ile Phe Tyr Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala  
 95 100 105 110  
 Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys  
 115 120 125  
 Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Met Glu  
 130 135 140  
 Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Pro Lys  
 145 150 155  
 Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Lys Asp Gly  
 160 165 170  
 Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp  
 175 180 185 190  
 Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala  
 195 200 205  
 Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Ile Leu Leu Glu  
 210 215 220  
 Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys  
 225 230 235

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 764 nucleotides

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

AAGCTTT	ATG	AGT	AAA	GGA	GAA	GAA	CTT	TTC	ACT	GGA	GTT	GTC	CCA	ATT	49
	Met	Ser	Lys	Gly	Glu	Glu	Leu	Phe	Thr	Gly	Val	Val	Pro	Ile	
CTT GTT	GAA	TTA	GAT	GGC	GAT	GTT	AAT	GGG	CAA	AAA	TTC	TCT	GTT	AGT	97
Leu Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly	Gln	Lys	Phe	Ser	Val	Ser	
GGA GAG	GGT	GAA	GGT	GAT	GCA	ACA	TAC	GGA	AAA	CTT	ACC	CTT	AAA	TTT	145
Gly Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly	Lys	Leu	Thr	Leu	Lys	Phe	
ATT TGC	ACT	ACT	GGG	AAG	CTA	CCT	GTT	CCA	TGG	CCA	ACG	CTT	GTC	ACT	193
Ile Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro	Trp	Pro	Thr	Leu	Val	Thr	
ACT CTC	ACT	TAT	GGT	GTT	CAA	TGC	TTT	TCT	AGA	TAC	CCA	GAT	CAT	ATG	241
Thr Leu	Thr	Tyr	Gly	Val	Gln	Cys	Phe	Ser	Arg	Tyr	Pro	Asp	His	Met	
AAA CAG	CAT	GAC	TTT	TTC	AAG	AGT	GCC	ATG	CCC	GAA	GGT	TAT	GTA	CAG	289
Lys Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met	Pro	Glu	Gly	Tyr	Val	Gln	

GAA AGA ACT ATA TTT TAC AAA GAT GAC GGG AAC TAC AAG ACA CGT GCT 337  
 Glu Arg Thr Ile Phe Tyr Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala

GAA GTC AAG TTT GAA GGT GAT ACC CTT GTT AAT AGA ATC GAG TTA AAA 385  
 Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys

GGT ATT GAT TTT AAA GAA GAT GGA AAC ATT CTT GGA CAC AAA ATG GAA 433  
 Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Met Glu

TAC AAT TAT AAC TCA CAT AAT GTA TAC ATC ATG GCA GAC AAA CCA AAG 481  
 Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Pro Lys

AAT GGC ATC AAA GTT AAC TTC AAA ATT AGA CAC AAC ATT AAA GAT GGA 529  
 Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Lys Asp Gly

AGC GTT CAA TTA GCA GAC CAT TAT CAA CAA AAT ACT CCA ATT GGC GAT 577  
 Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp

GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACG CAA TCT GCC 625  
 Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala

CTT TCC AAA GAT CCC AAC GAA AAG AGA GAT CAC ATG ATC CTT CTT GAG 673  
 Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Ile Leu Leu Glu

TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA CTA TAC AAA 721  
 Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys

TAA ATGTCCAGAC TTCCAATTGA CACTAAAGGG ATCCGAATTC 764

- (2) INFORMATION FOR SEQ ID NO:20:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 238 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: protein  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile  
                                   5                                  10

Leu Val Glu Leu Asp Gly Asp Val Asn Gly Gln Lys Phe Ser Val Ser  
   15                                  20                                  25                                  30

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe  
                                   35                                  40                                  45

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr  
                                   50                                  55                                  60

Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met  
                                   65                                  70                                  75

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln  
   80                                  85                                  90

Glu Arg Thr Ile Phe Tyr Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala  
   95                                  100                                  105                                  110

Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys  
                                   115                                  120                                  125

Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Met Glu  
                                   130                                  135                                  140

Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Pro Lys  
   145                                  150                                  155

Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Lys Asp Gly

000120:07E0T560

160		165		170
Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp				
175		180		185
				190
Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala				
	195		200	205
Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Ile Leu Leu Glu				
	210		215	220
Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys				
	225		230	235

000720-07E960